

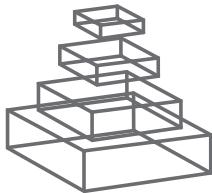
frontiers RESEARCH TOPICS

DNA POLYMERASES IN BIOTECHNOLOGY

Topic Editors
Andrew F. Gardner and Zvi Kelman



frontiers in
MICROBIOLOGY



FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2015
Frontiers Media SA.
All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-455-1

DOI 10.3389/978-2-88919-455-1

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

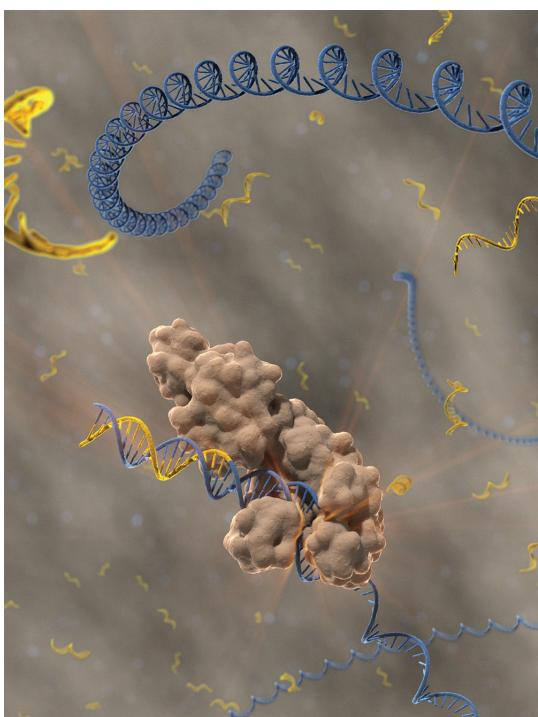
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

DNA POLYMERASES IN BIOTECHNOLOGY

Topic Editors:

Andrew F. Gardner, New England Biolabs, USA

Zvi Kelman, National Institute of Standards and Technology, USA



Polymerase image reprinted from www.neb.com (2014) with permission from New England Biolabs, Inc.

DNA polymerases are core tools for molecular biology including PCR, whole genome amplification, DNA sequencing and genotyping. Research has focused on discovery of novel DNA polymerases, characterization of DNA polymerase biochemistry and development of new replication assays. These studies have accelerated DNA polymerase engineering for biotechnology. For example, DNA polymerases have been engineered for increased speed and fidelity in PCR while lowering amplification sequence bias. Inhibitor resistant DNA polymerase variants enable PCR directly from tissue (i.e. blood). Design of DNA polymerases that efficiently incorporate modified nucleotide have been critical for development of next generation DNA sequencing, synthetic biology and other labeling and detection technologies. The Frontiers in Microbiology Research Topic on DNA polymerases in Biotechnology

aims to capture current research on DNA polymerases and their use in emerging technologies.

Table of Contents

- 05 DNA Polymerases in Biotechnology**
Andrew F. Gardner and Zvi Kelman
- 08 Evolution of Replicative DNA Polymerases in Archaea and their Contributions to the Eukaryotic Replication Machinery**
Kira S. Makarova, Mart Krupovic and Eugene V. Koonin
- 18 Structural Insights into Eukaryotic DNA Replication**
Sylvie Doublie and Karl E. Zahn
- 26 DNA Polymerases as Useful Reagents for Biotechnology – The History of Developmental Research in the Field**
Sonoko Ishino and Yoshizumi Ishino
- 34 DNA Polymerase Hybrids Derived From the Family-B Enzymes of Pyrococcus furiosus and Thermococcus kodakarensis: Improving Performance in the Polymerase Chain Reaction**
Ashraf M. Elshawadfy, Brian J. Keith, H'Ng Ee Ooi, Thomas Kinsman, Pauline Heslop and Bernard A. Connolly
- 48 Mutant Taq DNA Polymerases with Improved Elongation Ability as a Useful Reagent for Genetic Engineering**
Takeshi Yamagami, Sonoko Ishino, Yutaka Kawarabayasi and Yoshizumi Ishino
- 58 Replication Slippage of the Thermophilic DNA Polymerases B and D From the Euryarchaeota Pyrococcus abyssi**
Melissa Castillo-Lizardo, Ghislaine Henneke and Enrique Viguera
- 68 PCR Performance of a Thermostable Heterodimeric Archaeal DNA Polymerase**
Tom Killelea, Céline Ralec, Audrey Bossé and Ghislaine Henneke
- 79 Compartmentalized Self-Replication Under Fast PCR Cycling Conditions Yields Taq DNA Polymerase Mutants with Increased DNA-Binding Affinity and Blood Resistance**
Bahram Arezi, Nancy McKinney, Connie Hansen, Michelle Cayouette, Jeffrey Fox, Keith Chen, Jennifer Lapira, Sarah Hamilton and Holly Hogrefe
- 89 Bacteriophage T7 DNA Polymerase - Sequenase**
Bin Zhu
- 94 Engineering Processive DNA Polymerases with Maximum Benefit at Minimum Cost**
Linda J. Reha-Krantz, Sandra Woodgate and Myron F. Goodman

111 *DNA Polymerases Drive DNA Sequencing-by-Synthesis Technologies: Both Past and Present*

Cheng-Yao Chen

122 *DNA Polymerases Engineered by Directed Evolution to Incorporate Non- Standard Nucleotides*

Roberto Laos, J. Michael Thomson and Steven A. Benner

136 *A Novel Thermostable Polymerase for RNA and DNA Loop-Mediated Isothermal Amplification (LAMP)*

Yogesh Chander, Jim Koelbl, Jamie Puckett, Michael J. Moser, Audrey J. Klingele, Mark R. Liles, Abel Carrias, David A. Mead and Thomas W. Schoenfeld



DNA polymerases in biotechnology

Andrew F. Gardner^{1*} and Zvi Kelman^{2,3}

¹ New England Biolabs Inc., Ipswich, MA, USA

² National Institute of Standards and Technology, Gaithersburg, MD, USA

³ Institute for Bioscience and Biotechnology Research, Rockville, MD, USA

*Correspondence: gardner@neb.com

Edited by:

John R. Battista, Louisiana State University and A & M College, USA

Reviewed by:

Katarzyna Bebenek, National Institute of Environmental Health Sciences - National Institutes of Health, USA

Keywords: DNA polymerase, DNA polymerase evolution, DNA polymerase fidelity, DNA sequencing, molecular diagnostics, next generation sequencing, PCR, PCR inhibitors

Accurate duplication of parental DNA is a fundamental biological process, conserved in function across all life forms. All organisms depend on DNA polymerases for genome replication and maintenance. DNA polymerases also play central roles in modern molecular biology and biotechnology, enabling techniques including DNA cloning, the polymerase chain reaction (PCR), DNA sequencing, single nucleotide polymorphism (SNP) detection, whole genome amplification (WGA), synthetic biology, and molecular diagnostics. Each of these applications relies on the ability of polymerases to duplicate DNA, yielding a product that accurately represents the initial input. This book on "DNA Polymerases in Biotechnology" focuses on how detailed understanding of DNA polymerase structure and function informs protein engineering efforts, leading to development of novel reagents for molecular biology and clinical diagnostics.

DNA polymerases are classified into several families (A, B, C, D, X, Y) and reverse transcriptase (RT) based on primary amino acid sequence similarities (Burgers et al., 2001). The book leads off with several reviews that describe how these DNA polymerase families are evolutionarily (Makarova et al., 2014) and structurally (Doublie and Zahn, 2014) related as well as how polymerases have been utilized in biotechnology (Ishino and Ishino, 2014). Subsequent research articles build on this basic knowledge to describe how DNA polymerases are engineered as tools in biotechnology.

The best known and one of the earliest DNA polymerase-based biotechnology applications is PCR. Since its development over 30 years ago, PCR has been a foundational tool for amplifying and detecting specific alleles (Erlich et al., 1991). Advances in DNA polymerase fidelity, speed, and processivity continue to improve PCR workflows for genetic analysis, cloning, and diagnostics. Several articles in the issue highlight engineered polymerases with improved properties for PCR. Elshawadfy and co-workers demonstrate that combining desirable protein domains from several DNA polymerases into a single engineered chimeric enzyme can increase both speed and processivity during PCR (Elshawadfy et al., 2014). Similarly, Yamagami et al. create novel DNA polymerases by swapping domains from DNA polymerases found in hot springs to select for hybrid polymerases with desirable PCR properties (Yamagami et al., 2014). Castillo-Lizardo and colleagues analyze replication slippage during PCR

of repeat sequences and show that the processivity clamp of DNA polymerase, proliferating cell nuclear antigen (PCNA) (Indiani and O'Donnell, 2006), reduces slippage to permit error-free replication of repeat sequences (Castillo-Lizardo et al., 2014).

As nucleic acid analysis by PCR moves toward clinical diagnostics, there is a need for both faster DNA polymerases and those that are capable of directly amplifying DNA from clinical samples such as tissue, blood, body fluids, or stool to speed and simplify diagnostic workflows. Several papers characterize DNA polymerases that tolerate PCR inhibitors and allow rapid DNA amplification from clinical samples without DNA purification, thereby reducing analysis time, cost, and potential for contamination. The contribution by Killelea et al. demonstrates that a Family D polymerase from *Pyrococcus abyssi* is tolerant to high concentrations of PCR inhibitors while Arezi and colleagues describe a method to select for DNA polymerase variants that enable direct PCR from whole blood (Arezi et al., 2014; Killelea et al., 2014).

In addition to PCR, DNA polymerases play key roles in DNA sequencing technologies. Sanger DNA sequencing was used to sequence the first draft of the human genome in 2001 (Lander et al., 2001; Venter et al., 2001) and remains a standard and widespread method to determine DNA sequence. Several reviews describe the recent progress in the use of DNA polymerases for DNA sequencing. The review by Zhu examines the pivotal role of T7 DNA polymerase and its engineered derivatives in accelerating Sanger sequencing techniques (Zhu, 2014). Reha-Krantz et al. combine genetic and biochemical methods to identify T4 DNA polymerase mutants with increased processivity that along with T4 single stranded DNA binding protein (gp32) and T4 processivity factors (gp45 and gp44/62 complex) (Indiani and O'Donnell, 2006) improve Sanger sequencing of difficult DNA regions (Reha-Krantz et al., 2014). In the years since the human genome was first sequenced, new next generation sequencing methods have dramatically increased sequencing output while lowering costs (Mardis, 2011). Again, engineered DNA polymerases form the core of these next generation DNA sequencing-by-synthesis technologies. Chen reviews how DNA polymerases enable sequencing-by-synthesis technologies including the Illumina, Ion Torrent, and Pacific

Biosciences platforms (Chen, 2014). The contribution by Laos et al. details how DNA polymerases have been engineered to incorporate the modified nucleotides used in DNA sequencing, genotyping, and synthesis of artificial DNA (Laos et al., 2014).

Recently, point of care diagnostic tests that are cheap, reliable and do not depend on specialized instruments have emerged. For example, isothermal amplification techniques such as Loop-Mediated Amplification (LAMP) have been routinely used as diagnostic tests to detect infectious disease (Njiru, 2012). Chander and co-worker describe an engineered thermostable viral polymerase with RT and DNA polymerase activities that can be used in isothermal RT-LAMP detection of RNA (Chander et al., 2014). Additionally, they demonstrate that the reaction components can be lyophilized as a dry pellet to allow storage without refrigeration and may be used in the field as a simple diagnostic test for RNA viruses.

FUTURE CHALLENGES

Engineered DNA polymerases will continue to play important roles in biotechnology and the delivery of health care. Over the next several years, molecular methods that are easier, cheaper, and faster will emerge. At the same time, molecular biology will move toward analysis of low concentration biomolecules (i.e., a single set of chromosomes). Unfortunately, tools for analysis of minute quantities of DNA are currently inadequate or technically challenging. For example, advances in sequencing technology (i.e., nanopore sequencing) can use extremely long DNA but methods to create long DNAs have not kept pace (Branton et al., 2008; Metzker, 2010; Shendure et al., 2011). Novel amplification techniques are also required to profile genetic variations among single cells (Navin and Hicks, 2011; Schubert, 2011) because the quantity of genomic DNA from a single cell is insufficient to sequence directly. Therefore, DNA must first be amplified prior to further analysis (Kalisky and Quake, 2011; Kalisky et al., 2011).

Additionally, synthetic biology aims to design new biological systems such as genetic pathways, operons, and genomes (Montague et al., 2012) and thus may require long, chromosome-size, amplification. Pathway engineering relies on assembling the DNA coding for the desired characteristics and then using a host to activate the pathway *in vivo*. Current methods for DNA assembly are limited to about 20 kb and larger fragments must be assembled *in vivo* at a very low frequency thereby limiting utility. Furthermore, current DNA polymerases introduce errors during amplification and thus DNA polymerases with very low error rates are needed to ensure that long, amplified DNA are exact copies of the starting material.

Therefore, novel DNA amplification systems are needed to accelerate progress in emerging technologies and to make high-fidelity *in vitro* genome analysis and manipulation routine. Engineered DNA polymerases or cellular replication machineries capable of amplifying large DNA fragments have the potential to enable single cell genomics, genome synthesis, and manipulation. This issue summarizes the known properties of various DNA polymerase systems and how DNA polymerases are currently being manipulated to meet these growing demands.

REFERENCES

- Arezi, B., McKinney, N., Hansen, C., Cayouette, M., Fox, J., Chen, K., et al. (2014). Compartmentalized self-replication under fast PCR cycling conditions yields Taq DNA polymerase mutants with increased DNA-binding affinity and blood resistance. *Front. Microbiol.* 5:408. doi: 10.3389/fmicb.2014.00408
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., et al. (2008). The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* 26, 1146–1153. doi: 10.1038/nbt.1495
- Burgers, P. M., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., et al. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.* 276, 43487–43490. doi: 10.1074/jbc.R100056200
- Castillo-Lizardo, M., Henneke, G., and Viguera, E. (2014). Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota *Pyrococcus abyssi*. *Front. Microbiol.* 5:403. doi: 10.3389/fmicb.2014.00403
- Chander, Y., Koelbl, J., Puckett, J., Moser, M. J., Klingele, A. J., Liles, M. R., et al. (2014). A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.* 5:395. doi: 10.3389/fmicb.2014.00395
- Chen, C. Y. (2014). DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front. Microbiol.* 5:305. doi: 10.3389/fmicb.2014.00305
- Doublie, S., and Zahn, K. E. (2014). Structural insights into eukaryotic DNA replication. *Front. Microbiol.* 5:444. doi: 10.3389/fmicb.2014.00444
- Elshawadfy, A. M., Keith, B. J., Ee Ooi, H., Kinsman, T., Heslop, P., and Connolly, B. A. (2014). DNA polymerase hybrids derived from the family-B enzymes of *Pyrococcus furiosus* and *Thermococcus kodakarensis*: improving performance in the polymerase chain reaction. *Front. Microbiol.* 5:224. doi: 10.3389/fmicb.2014.00224
- Erlich, H. A., Gelfand, D., and Sninsky, J. J. (1991). Recent advances in the polymerase chain reaction. *Science* 252, 1643–1651. doi: 10.1126/science.2047872
- Indiani, C., and O'Donnell, M. (2006). The replication clamp-loading machine at work in the three domains of life. *Nat. Rev. Mol. Cell Biol.* 7, 751–761. doi: 10.1038/nrm2022
- Ishino, S., and Ishino, Y. (2014). DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Front. Microbiol.* 5:465. doi: 10.3389/fmicb.2014.00465
- Kalisky, T., Blainey, P., and Quake, S. R. (2011). Genomic analysis at the single-cell level. *Annu. Rev. Genet.* 45, 431–445. doi: 10.1146/annurev-genet-102209-163607
- Kalisky, T., and Quake, S. R. (2011). Single-cell genomics. *Nat. Methods* 8, 311–314. doi: 10.1038/nmeth0411-311
- Killelea, T., Ralec, C., Bosse, A., and Henneke, G. (2014). PCR performance of a thermostable heterodimeric archaeal DNA polymerase. *Front. Microbiol.* 5:195. doi: 10.3389/fmicb.2014.00195
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Laos, R., Thomson, J. M., and Benner, S. A. (2014). DNA polymerases engineered by directed evolution to incorporate nonstandard nucleotides. *Front. Microbiol.* 5:565. doi: 10.3389/fmicb.2014.00565
- Makarova, K. S., Krupovic, M., and Koonin, E. V. (2014). Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front. Microbiol.* 5:354. doi: 10.3389/fmicb.2014.00354
- Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature* 470, 198–203. doi: 10.1038/nature09796
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Montague, M. G., Lartigue, C., and Vashee, S. (2012). Synthetic genomics: potential and limitations. *Curr. Opin. Biotechnol.* 23, 659–665. doi: 10.1016/j.copbio.2012.01.014
- Navin, N., and Hicks, J. (2011). Future medical applications of single-cell sequencing in cancer. *Genome Med.* 3, 31. doi: 10.1186/gm247
- Njiru, Z. K. (2012). Loop-mediated isothermal amplification technology: towards point of care diagnostics. *PLoS Negl. Trop. Dis.* 6:e1572. doi: 10.1371/journal.pntd.0001572

- Reha-Krantz, L. J., Woodgate, S., and Goodman, M. F. (2014). Engineering processive DNA polymerases with maximum benefit at minimum cost. *Front. Microbiol.* 5:380. doi: 10.3389/fmicb.2014.00380
- Schubert, C. (2011). Single-cell analysis: the deepest differences. *Nature* 480, 133–137. doi: 10.1038/480133a
- Shendure, J. A., Porreca, G. J., Church, G. M., Gardner, A. F., Hendrickson, C. L., Kieleszawa, J., et al. (2011). Overview of DNA sequencing strategies. *Curr. Protoc. Mol. Biol.* (edited by Frederick M Ausubel [et al.]) Chapter 7, Unit 7.1. doi: 10.1002/0471142727.mb0701s96
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040
- Yamagami, T., Ishino, S., Kawarabayashi, Y., and Ishino, Y. (2014). Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering. *Front. Microbiol.* 5:461. doi: 10.3389/fmicb.2014.00461
- Zhu, B. (2014). Bacteriophage T7 DNA polymerase - Sequenase. *Front. Microbiol.* 5:181. doi: 10.3389/fmicb.2014.00181

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 October 2014; accepted: 13 November 2014; published online: 01 December 2014.

Citation: Gardner AF and Kelman Z (2014) DNA polymerases in biotechnology. *Front. Microbiol.* 5:659. doi: 10.3389/fmicb.2014.00659

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Gardner and Kelman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery

Kira S. Makarova¹, Mart Krupovic² and Eugene V. Koonin^{1*}

¹ National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

² Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, Paris, France

Edited by:

Zvi Kelman, University of Maryland,
USA

Reviewed by:

Thijs Ettema, Uppsala University,
Sweden

Uri Gophna, Tel Aviv University,
Israel

***Correspondence:**

Eugene V. Koonin, National Center
for Biotechnology Information,
National Library of Medicine,
National Institutes of Health,
Building 38A, Room 5N503, 8600
Rockville Pike, Bethesda, MD
20894, USA
e-mail: koonin@ncbi.nlm.nih.gov

The elaborate eukaryotic DNA replication machinery evolved from the archaeal ancestors that themselves show considerable complexity. Here we discuss the comparative genomic and phylogenetic analysis of the core replication enzymes, the DNA polymerases, in archaea and their relationships with the eukaryotic polymerases. In archaea, there are three groups of family B DNA polymerases, historically known as PolB1, PolB2 and PolB3. All three groups appear to descend from the last common ancestors of the extant archaea but their subsequent evolutionary trajectories seem to have been widely different. Although PolB3 is present in all archaea, with the exception of Thaumarchaeota, and appears to be directly involved in lagging strand replication, the evolution of this gene does not follow the archaeal phylogeny, conceivably due to multiple horizontal transfers and/or dramatic differences in evolutionary rates. In contrast, PolB1 is missing in Euryarchaeota but otherwise seems to have evolved vertically. The third archaeal group of family B polymerases, PolB2, includes primarily proteins in which the catalytic centers of the polymerase and exonuclease domains are disrupted and accordingly the enzymes appear to be inactivated. The members of the PolB2 group are scattered across archaea and might be involved in repair or regulation of replication along with inactivated members of the RadA family ATPases and an additional, uncharacterized protein that are encoded within the same predicted operon. In addition to the family B polymerases, all archaea, with the exception of the Crenarchaeota, encode enzymes of a distinct family D the origin of which is unclear. We examine multiple considerations that appear compatible with the possibility that family D polymerases are highly derived homologs of family B. The eukaryotic DNA polymerases show a highly complex relationship with their archaeal ancestors including contributions of proteins and domains from both the family B and the family D archaeal polymerases.

Keywords: DNA replication, archaea, mobile genetic elements, DNA polymerases, enzyme inactivation

INTRODUCTION

Recent experimental and comparative genomic studies on DNA replication systems have revealed their remarkable plasticity in each of the three domains of cellular life (Li et al., 2013; Makarova and Koonin, 2013; Raymann et al., 2014). In particular, archaea, members of the prokaryotic domain that gave rise to the information processing systems of eukaryotes, show remarkable diversity even with respect to the core components of the replication machinery, the DNA polymerases (DNAPs) (Makarova and Koonin, 2013). The main replicative polymerases of archaea belong to the B family of Palm domain DNAPs (Burgers et al., 2001) which is also widely represented in eukaryotes, eukaryotic and bacterial viruses, as well as some bacteria; however, in bacteria, these polymerases appear to be of viral origin and are involved mainly in repair whereas replication relies on a distinct, unrelated enzyme (Gawel et al., 2008). In addition to the polymerase core, which consists of three domains known as palm, fingers and thumb, most of the B family DNAPs contain an N-terminal 3'-5' exonuclease domain and a uracil-recognition domain (Hopfner

et al., 1999; Steitz and Yin, 2004; Rothwell and Waksman, 2005; Delagoutte, 2012).

Family B DNAPs are present in all archaeal lineages, and many archaea have multiple paralogs some of which appear to be inactivated; at least two paralogs can be traced to the Last Archaeal Common Ancestor (LACA) (Rogozin et al., 2008; Makarova and Koonin, 2013). In addition to the archaeal chromosomes, family B DNAPs are encoded by several mobile genetic elements (MGEs) that replicate in archaeal cells and could contribute to horizontal transfer of DNAPs (Filee et al., 2002). In particular, family B DNAPs closely related to those found in the host species are encoded by haloarchaeal head-tailed viruses such as *Halorubrum* myoviruses HF1, HF2 (Filee et al., 2002; Tang et al., 2002) and HSTV-2 (Pietila et al., 2013) whereas more diverged protein-primed Family B DNAPs have been identified in other haloviruses such as His1 and His2 (Bath et al., 2006). Furthermore, recently, family B DNAPs have been identified in a new group of self-synthesizing mobile elements, called casposons because they apparently employ Cas1, originally known

as a component of the CRISPR-Cas immunity systems, as their integrase (Makarova et al., 2013; Krupovic et al., 2014a).

In addition to the family B polymerases, most of the archaeal lineages, with the exception of the Crenarchaeota, encode the unique family D DNAP (Cann et al., 1998) that accordingly can be inferred to have been present in LACA. The family D polymerases consist of two subunits. The large subunit DP2 is a multidomain protein which forms a homodimer that is responsible for the polymerase activity (Shen et al., 2001; Matsui et al., 2011). The DP2 protein does not show significant sequence similarity with any proteins except for the two C-terminal Zn finger domains. The structure of the complete DP2 protein so far has not been solved but the structure of the N-terminal domain reveals a unique fold (Matsui et al., 2011). The small subunit DP1 contains at least two domains, an ssDNA-binding OB-fold, and a 3'-5' exonuclease domain of the metallophosphatase MPP family. The DP1 protein is the ancestor of the small B subunits of eukaryotic replicative DNAPs that, however, have lost the catalytic amino acid residues of the 3'-5' exonuclease (Aravind and Koonin, 1998; Klinge et al., 2009). Evidence has been presented that in euryarchaea the family D DNAP specializes in the synthesis of the lagging strand whereas the family B DNAP, PolB3, is involved in the leading strand synthesis (Henneke et al., 2005). However, at least in *Thermococcus kodakarensis*, the family D DNAP is sufficient for the replication of both strands (Cubanova et al., 2013). The Crenarchaeota lack the family D DNAP but possess at least one additional active DNAP of the B family, suggesting that the two distinct B family DNAPs specialize in the leading and lagging strand replication, respectively, as is the case in eukaryotes. In particular, biochemical data suggest that in *Sulfolobus solfataricus*, one family B polymerase (PolB1/Dpo1) is responsible for the synthesis of the leading strand whereas the other one, PolB3/Dpo3, is involved in the synthesis of the lagging strand (Bauer et al., 2012).

Some crenarchaeal and euryarchaeal plasmids encode palm domain polymerases of the archaeo-eukaryotic primase superfamily (Iyer et al., 2005), known as prim-pol, but in these plasmids the protein apparently is employed for initiation of replication rather than elongation (Iyer et al., 2005; Lipps, 2011; Krupovic et al., 2013; Gill et al., 2014).

Here we summarize the results of an updated comparative genomic and phylogenetic analysis of archaeal polymerases, focusing primarily on the diversity of Family B, including the polymerases associated with proviruses and mobile elements, and discuss their evolutionary relationships with eukaryotic DNAPs.

COMPARATIVE GENOMIC AND PHYLOGENETIC ANALYSIS OF ARCHAEL DNA POLYMERASES

PHYLOGENY, DOMAIN ARCHITECTURE AND GENE NEIGHBORHOODS OF B FAMILY DNAPs IN ARCHAEA

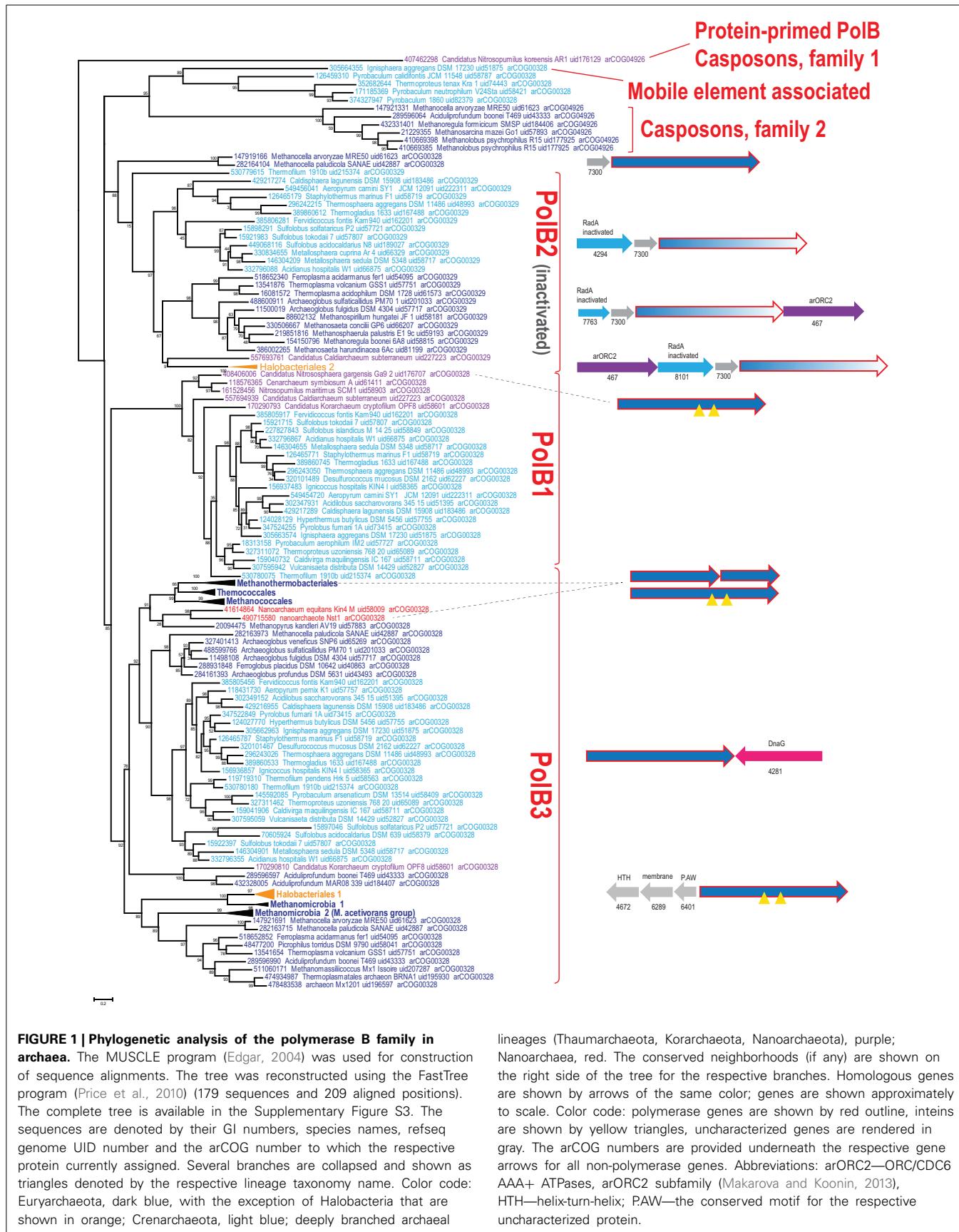
Using the latest recent update of archaeal clusters of orthologous genes (arCOGs) (Wolf et al., 2012) which includes 168 complete genome sequences of archaea (Refseq update as of February 2014), we reconstructed a phylogenetic tree of family B polymerases for a representative set of archaeal genomes and analyzed their gene context (Figure 1). One of the selected sequences (YP_006773615 from *Candidatus Nitrosopumilus koreensis*) belongs to the distinct, protein-primed DNAP family (see

discussion below) and thus was used as an outgroup (Figure 1). Another protein (YP_007906966 from *Archaeoglobus sulfatocalidus*) is extremely diverged and poorly alignable and therefore has not been included in the tree reconstruction. Consistent with previous observations (Edgell et al., 1998; Rogozin et al., 2008), the tree encompassed three large branches: (i) PolB3, the “major” DNAP, present in all archaea except Thaumarchaeota, (ii) PolB1, the “minor” DNAP, present only in the TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota) superphylum (Guy and Ettema, 2011; Martijn and Ettema, 2013) and (iii) PolB2, a distinct family of DNAP homologs most of which appear to be inactivated as inferred from the replacement of the catalytic amino acid residues (Rogozin et al., 2008) and show a patchy distribution in most archaeal lineages (Figures 1, 2, Supplementary Table S1).

Despite the presence in most archaeal genomes, the PolB3 branch shows little topological congruence with the archaeal phylogeny that was established primarily through phylogenetic analysis of multiple translation, transcription and replication system components (Guy and Ettema, 2011; Yutin et al., 2012; Podar et al., 2013; Raymann et al., 2014). The deviations include the polyphyly of Euryarchaeota, Methanomicrobia, and Thermoplasmatales, and paraphyly of Sulfolobales-Desulfurococcales with respect to Thermoproteales. These discrepancies suggest that the history of archaeal Family B DNAPs included multiple horizontal gene transfer (HGT) events and/or major accelerations of evolution. No recent duplications are observed within this group of polymerases but some archaea possess two versions of PolB3 that could have different origins. In particular, acquisition of two versions of PolB3 (one from Archaeoglobales and another from Thermoplasmatales), followed by the loss of the ancestral methanomicrobial gene, seems likely for the genus *Methanocella*.

Several groups of archaea contain intein insertions in the PolB3 gene, up to three per gene (Perler, 2002). Inteins are parasitic genetic elements that insert into protein-coding genes, perform self-splicing at the protein level and typically encode an endonuclease that mediates intein gene propagation into ectopic DNA sites (Perler et al., 1994; Gogarten et al., 2002). The majority of intein insertion sites in PolB3 genes are shared between different archaea but some are lineage-specific (Perler, 2002; MacNeill, 2009). It appears likely that the split PolB3 genes in Methanobacteriales (Kelman et al., 1999) evolved as a result of erratic intein excision, especially considering that in the tree these split DNAP genes cluster with Methanococcales and Thermococcales which both contain inteins in PolB3 genes (Figure 1). Similarly, a split PolB gene, in this case with the two parts non-adjacent, is found in *Nanoarchaeum equitans* where it could be trans-spliced via an intein parts of which are associated with the two split gene fragments (Perler, 2002; Choi et al., 2006). In the recently sequenced nanoarchaeon Nst1, the orthologous PolB3 gene is not split (Podar et al., 2013), suggesting that intein insertion and split occurred late in the evolution of the Nanoarchaeota.

In most of the archaea, PolB3 genes do not form conserved genomic neighborhoods. The only notable exception is a conserved genomic context of this gene in most crenarchaea that



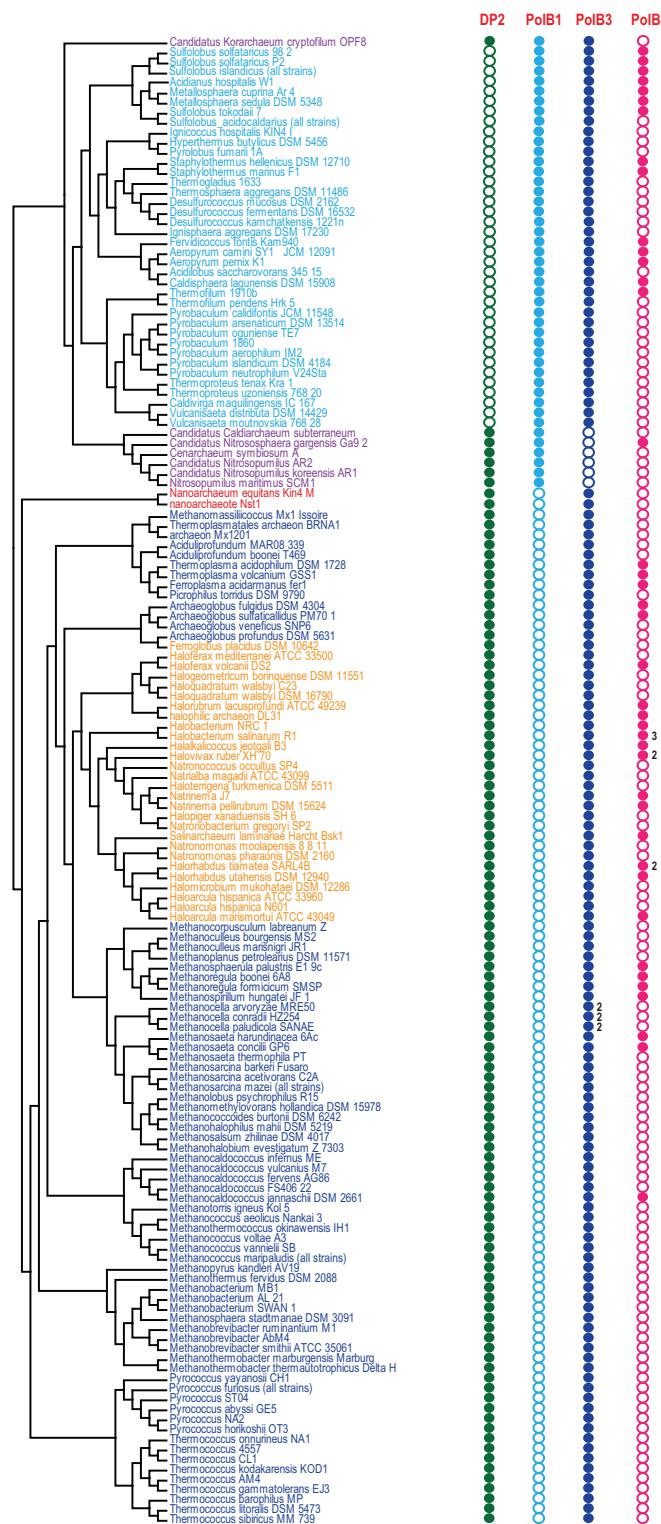


FIGURE 2 | Phyletic patterns for the major groups of archaeal DNAPs. Phyletic patterns for the indicated protein families (filled circles show presence and empty circles show absence) are superimposed over the phylogenetic tree of archaea. The number

of paralogs are indicated by a number, if more than one paralog of the respective family is encoded in the genome. The tree topology is based on the phylogeny of concatenated ribosomal proteins.

includes the bacterial-type DNA primase *dnaG*; however, the *polB3* and *dnaG* genes are oriented convergently and accordingly are transcribed from different promoters. In addition, in all haloarchaeal genomes, PolB3 might be co-regulated with three uncharacterized genes that are specific to this group of archaea (e.g., HVO_0855-HVO_0857 from *Haloferax volcanii*); the protein product of one of these genes (HVO_0855) contains a helix-turn-helix DNA-binding domain, suggesting that it could be a regulator of PolB3 transcription (**Figure 1**).

The second major branch of archaeal family B DNAPs includes the replicative polymerases of the PolB1 group that is represented in all members of the TACK superphylum (**Figure 2**). Most of the Thaumarchaeota possess only this form of active family B DNAP whereas Korarchaeum and Crenarchaeota encode both PolB3 and PolB1. In a striking contrast to PolB3, the topology of this branch is almost fully consistent with the archaeal phylogeny, indicative of a primarily vertical mode of evolution of this gene. So far only in *Nitrososphaera gargensis*, two inteins are inserted into the PolB1 gene (**Figure 1**).

The third large group of family B DNAPs that includes the experimentally characterized PolB2/Dpo2 of *S. solfataricus* shows a patchy distribution in archaea but is rapidly growing with the sequencing of new genomes that have been found to encompass this gene, along with several bacteria (Rogozin et al., 2008). The PolB2 family is currently represented in Crenarchaeota, Euryarchaea and also in *Caldarchaeum subterraneum*, the only known member of the putative phylum Aigarchaeota (**Figure 1**). The topology of this branch is generally consistent with a predominantly vertical mode of evolution, along with multiple losses in several archaeal lineages. It appears likely that in the case of this group, the deviations from the archaeal phylogeny are due primarily to increased rates of evolution of this gene in some lineages (**Figure 1**). Thus, along with the PolB3 and PolB1 groups, PolB2 probably was already represented in LACA. Sequence comparison of this subfamily with other family B DNAPs shows that, in most members, multiple catalytic residues of both the polymerase and the exonuclease domains are replaced, suggesting that these proteins are inactivated DNAPs (Rogozin et al., 2008). However, very weak activities of both enzymatic domains have been reported for a single member of this group, PolB2/Dpo2 of *S. solfataricus* (Choi et al., 2011).

Recent comparative genomic analysis identified an association between PolB2 genes, an uncharacterized gene of arCOG07300 and a *radA*-like gene in Sulfolobales (Makarova and Koonin, 2013). We analyzed the genomic neighborhoods for this family in greater detail and found that many diverged members of arCOG07300 have been missed originally due to the low sequence similarity with proteins from Sulfolobales but were now detected by using more sensitive methods, such as PSIBLAST, allowing to expand the family considerably (Supplementary Figure S1). The arCOG07300 proteins are small (~90 aa), alpha-helical proteins that do not show statistically significant similarity with any available protein sequences. Three arCOGs (arCOG07763, arCOG04294, arCOG08101) in the predicted operons with the inactivated polymerase PolB2 and arCOG07300 belong to the RadA family but all appear to be inactivated as judged by the substitution of the key amino acid residues implicated in ATP

binding and hydrolysis (Supplementary Figure S2). In one of these proteins (arCOG07763), the P-loop ATPase domains deteriorated so severely that similarity to RadA could be detected only using such sensitive methods as HHpred (Supplementary Figure S2). Because the phyletic patterns of arCOG07763, arCOG04294, arCOG08101 are complementary and the respective genes are embedded in the same genomic context, these genes appear to be orthologs that have evolved at high rates, losing readily detectable sequence similarity. Several haloarchaea possess an additional copy of a two gene operon that consists of arCOG07763 and arCOG07300. In two *Methanocella* species, the arCOG07300 gene is also present in the same operon with predicted active DNAPs which form the sister group to the inactivated PolB2/Dpo2 (**Figure 1**), suggesting that the functional link with arCOG07300 evolved before polymerase inactivation. In many euryarchaeal genomes, the neighborhood also includes an arORC2 family gene (Makarova and Koonin, 2013), an ATPase component of origin recognition complex (**Figure 1**). The same three gene families are linked also in the several bacterial genomes that encode a PolB2 homolog. In addition to these three genes, in some bacteria, *lexA*, the SOS response master repressor gene, is located in the same predicted operon. This association implies that the putative protein complex encoded by this operon is involved in DNA damage response. A typical example of these associations is a locus in *Leptospirillum ferriphilum* that consists of four genes LFML04_0990-LFML04_0993 encoding, respectively, an “inactivated” polymerase, a homolog of arCOG07300, inactivated *radA* and *lexA*. Taken together, these observations indicate that PolB2/Dpo2, inactivated RadA and arCOG07300 proteins are most likely functionally linked and could also form a complex given that proteins encoded in evolutionarily conserved operons often interact both physically and functionally (Dandekar et al., 1998; Quax et al., 2013).

Given its wide spread and likely ancestral provenance in archaea, this complex might perform important, albeit dispensable function in DNA damage repair, more specifically, perhaps in the elimination of stalled replication forks, and/or in the regulation of DNA replication.

The presence of the arORC2 family gene, which encodes an ATPase component of the origin recognition complex, in the same neighborhood of many euryarchaeal genomes implies a replication-related function (Makarova and Koonin, 2013) (**Figure 1**). Recently, it has been shown that in *Haloferax volcanii* a RadA protein is required for initiation of replication in origin-less cells (Hawkins et al., 2013). Although the RadA shown to be involved in this process is an active ATPase and belongs to a different family (arCOG00417), given the association of the putative PolB2-inactive RadA-arCOG07300 operon with the arORC2 gene, the complex of these proteins might be involved in an alternative mechanism of replication initiation or in the regulation of origin recognition. Clearly, an important aspect of the further characterization of this predicted complex is the determination of the presence or absence (as suggested by comparative sequence analysis) of enzymatic activities in PolB2.

Several archaea possess another, divergent B family DNAP (arCOG04926) that is predicted to be active. Recently, it has been shown that this gene is tightly associated with several other

genes, including Cas1 (a CRISPR-Cas system gene), and belongs to a new class of mobile elements called Casposons (see details below). A sister branch of this family includes active polymerases from several closely related genomes of Thermoproteales and a single representative of Desulfurococcales, *Ignisphaera aggregans* (Figure 1). In *I. aggregans*, the DNAP gene of this group probably belongs to a provirus (see below) whereas the respective genes in Thermoproteales do not display any conserved genomic associations and are unlikely to belong to mobile genetic elements although their origin from such elements cannot be ruled out.

DNA POLYMERASES ENCODED WITHIN INTEGRATED MOBILE ELEMENTS

Mobile genetic elements (MGE), such as viruses and plasmids, often encode their own genome replication proteins. In archaea, viruses from at least four different families are known to encode DNA polymerases. Tailed viruses of the order *Caudovirales* encode RNA-primed family B DNA polymerases (PolB) (Sencilo et al., 2013), whereas certain members of the families *Ampullaviridae* (Peng et al., 2007), *Fuselloviridae* (Bath et al., 2006; Krupovic et al., 2014b) and *Pleolipoviridae* (Bath et al., 2006; Pietila et al., 2012) carry genes for protein-primed PolBs. Integration of MGEs that contain genes for cellular-like replication proteins into the host chromosome can be and often is confused with the duplication of the *bona fide* cellular genes encoding these proteins (Krupovic et al., 2010; Forterre and Prangishvili, 2013). Therefore, careful gene neighborhood analysis is necessary to ascertain the provenance of replication protein genes in genomes of cellular organisms, especially when multiple paralogs of a given gene appear to be present.

With regard to DNAPs, two types of elements encoding diverse family B polymerases are integrated in the genomes of diverse archaea (Figure 3). The first group includes the recently discovered transposon-like elements called Casposons

(Krupovic et al., 2014a). Unlike other known mobile genetic elements, casposons apparently rely on Cas1 endonucleases, the key enzymes of the prokaryotic CRISPR-Cas immunity (hence the name), for integration into the cellular genome. These elements are found in both bacteria and archaea. Casposons are 7–20 kb in length and are surrounded by terminal inverted repeats and target site duplications (Figure 3). Three families of casposons have been defined based on the phylogenetic analysis of the Cas1 endonucleases, gene content and taxonomic distribution. Family 1 casposons are thus far exclusively found in Thaumarchaeota (4 elements) and encode protein-primed PolBs that are most closely related to the corresponding proteins of archaeal viruses His1 (*Fuselloviridae*) and His2 (*Pleolipoviridae*). Phylogenetic analysis of the viral and casposon pPolB suggests that there has been exchange of the pPolB genes between these two types of MGEs (Krupovic et al., 2014a).

Casposons of families 2 and 3 encode typical RNA-primed PolBs and are respectively found in the genomes of euryarchaeota (11 casposons) and bacteria (4 casposons). In the phylogenetic tree of PolB, these bacterial and archaeal casposons form a clade that emerges as a sister group to the DNAPs of different species of the crenarchaeal class Thermoprotei (Figure 1). Notably, in the latter group, PolB of *Ignisphaera aggregans* DSM 17230 is also encoded within an integrated mobile element which is, however, unrelated to the casposons (see below).

The second type of PolB-encoding MGEs includes two elements, IgnAgg-E3 (24.7 kb) and ArcSul-E1 (21.2 kb), found in the genomes of the crenarchaeon *I. aggregans* and the euryarchaeon *Archaeoglobus sulfaticallidus* PM70-1, respectively (Figure 3). These two elements share genes neither with each other nor with known archaeal viruses or plasmids (a detailed description of IgnAgg-E3 and ArcSul-E1 will be published elsewhere) and accordingly could be founding members of two new groups of MGEs.

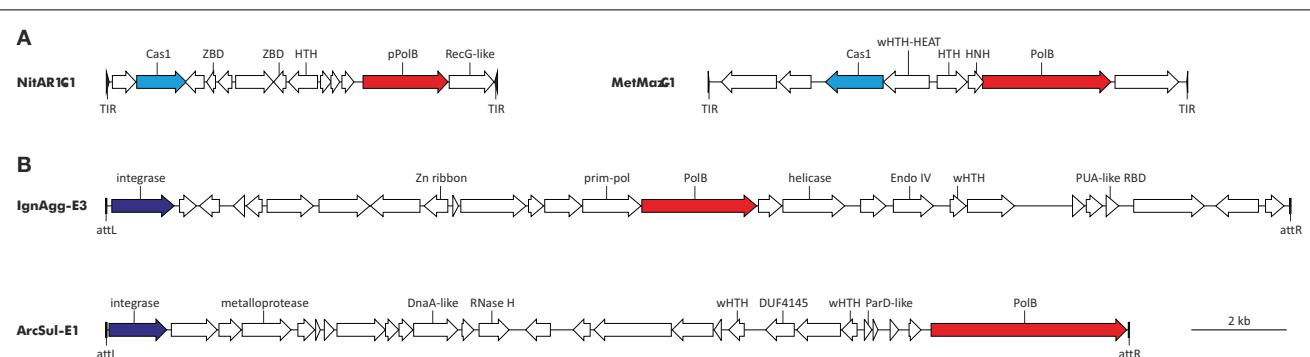


FIGURE 3 | Genome maps of archaeal PolB-encoding mobile genetic elements. (A) Casposons of families 1 and 2. NitAR1-C1 is present in the genome of *Candidatus Nitrosopumilus koreensis* AR1 (NC_018655; nucleotide coordinates: 655308 to 663492), whereas MetMaz-C1 is from *Methanoscincina mazei* Go1 (NC_003901; nucleotide coordinates: 3946601 to 3956653). (B) Tyrosine recombinases-encoding elements. IgnAgg-E3 is found in the genome of *Ignisphaera aggregans* DSM 17230 (NC_014471; nucleotide coordinates: 1844012 to 1868704) and ArcSul-E1 is from

Archaeoglobus sulfaticallidus PM70-1 (NC_021169; nucleotide coordinates: 873590 to 894826). Predicted protein-coding genes are indicated with arrows, indicating the direction of transcription. Genes for PolBs are shown in red, cas1 genes are in cyan, and genes for tyrosine recombinases are colored blue. Abbreviations: TIR, terminal inverted repeats; att, attachment site; ZBD, Zinc-binding domain-containing protein; HNH, HNH family endonuclease; (w)HTH, (winged) helix-turn-helix proteins; RBD, RNA-binding domain.

EVOLUTIONARY RELATIONSHIPS OF ARCHAEOAL AND EUKARYOTIC DNA POLYMERASES

Based on the above considerations and the respective phyletic patterns, three family B polymerases, PolB1, PolB2 ("inactivated" DNAPs) and PolB3, could be projected to LACA. In addition to these three families, two subunits of family D polymerase, arCOG04455 and arCOG04447, and a family Y polymerase, arCOG04582 (see the respective phyletic patterns in the Supplementary Table S1) also are likely to be ancestral. A polymerase of family X, although common in archaea, cannot be projected to LACA with confidence (Wolf et al., 2012). The latter two polymerases (families X and Y) are unlikely to be involved in genome replication. In bacteria and eukaryotes, members of both families have been thoroughly characterized and shown to function in DNA repair (Jarosz et al., 2007; Moon et al., 2007; Silverstein et al., 2010; Sharma et al., 2013). The experimental data suggests that PolB1 and the family D DNAP are the main replicative polymerases in crenarchaea and euryarchaea, respectively, whereas PolB3 appears to be involved in the replication of the lagging strand in most archaea (Cubanova et al., 2013).

Most eukaryotes possess four paralogous family B DNAPs denoted Pol- α , Pol- δ , Pol- ϵ , and Pol- ζ , four family Y polymerases (Yang, 2014), four family X polymerases (Bebenek et al., 2014) and two family A polymerases involved in mitochondrial

replication and DNA repair (Burgers et al., 2001). All these polymerases seem to have been present in the last eukaryotic common ancestor (LECA). The functions of family B polymerases in eukaryotes are diversified: Pol- ϵ is the main replicative polymerase specialized in the replication of the leading strand, Pol- δ replicates the lagging strand, Pol- α is the main component of the eukaryote-specific primase complex, which synthesizes short DNA primers during the lagging strand replication (Kunkel and Burgers, 2008; Pavlov and Shcherbakova, 2010 and references therein), and Pol- ζ is involved in lesion bypass (Sharma et al., 2013). Furthermore, the functions of all family B DNAPs in eukaryotes require an additional small subunit, the same for all family B DNAPs (Bell and Dutta, 2002).

Domain architectures and the relationships between archaeal and eukaryotic replicative polymerase families are schematically shown in **Figure 4A**. The small subunits evolved from the small subunit (DP1) of the archaeal family D polymerase which in archaea is an 3'-5'-exonuclease of the MPP superfamily that appears to be involved in proofreading during archaeal DNA replication (Aravind and Koonin, 1998; Jokela et al., 2004). However, the homologous small subunit of the eukaryotic DNAPs has lost the catalytic amino acid residues and performs an architectural role in the DNAP complex (Aravind and Koonin, 1998; Klinge et al., 2009; Yamasaki et al., 2010).

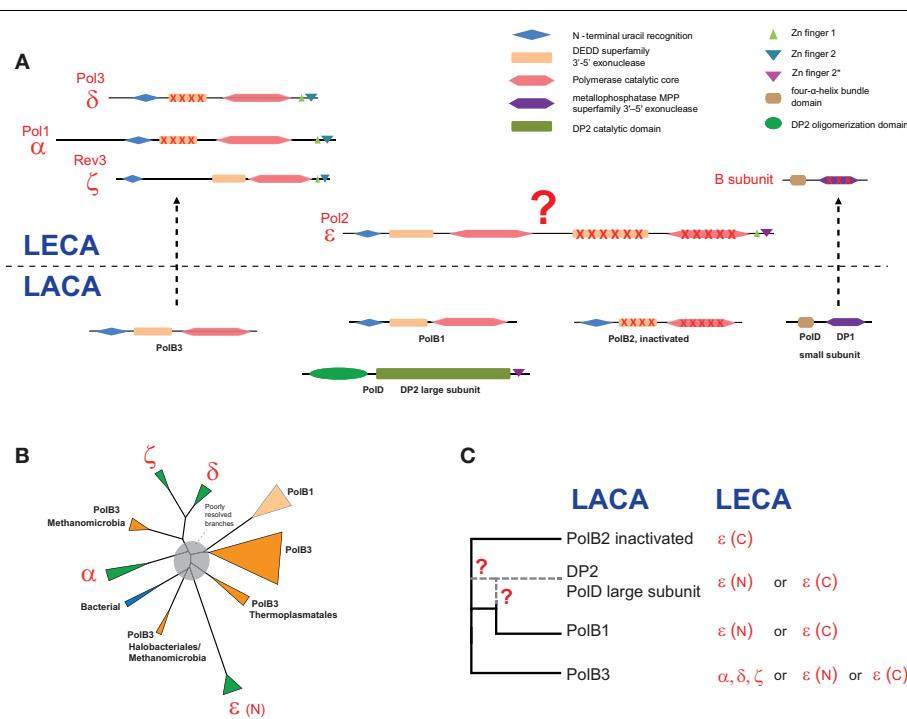


FIGURE 4 | Reconstruction of the complements of replicative DNAPs in the last archaeal and eukaryotic ancestors and a hypothetical scenario of their evolutionary relationships. (A) Polymerase B and D family genes projected to archaeal (LACA) and eukaryotic (LECA) last common ancestors and their domain organization. Homologous domains are shown by shapes of the same color. Inactivated domains are crossed. For eukaryotic polymerase families, human and yeast gene names are provided. **(B)** The unrooted phylogenetic tree of active polymerases of B family. The MUSCLE program

(Edgar, 2004) was used for construction of multiple sequence alignments. The tree was reconstructed using the FastTree program (Price et al., 2010) (141 sequences and 264 aligned positions). The complete tree is available in the Supplementary Figure S4. The tree is rendered as a scheme, with all major groups collapsed. **(C)** The inferred evolutionary relationships between archaeal and eukaryotic replicative DNAPs. Details on the involvement of PolD in the evolution of eukaryotic DNAPs are discussed in the text. The question mark denotes an uncertainty in evolutionary scenario.

The evolutionary relationships between the polymerase sub-units themselves are much more difficult to establish due to the multiplicity of paralogs in both archaea and eukaryotes, and the apparent differences in the evolutionary rates resulting in poorly resolved phylogenetic trees (Edgell et al., 1998; Filee et al., 2002; Tahirov et al., 2009). Furthermore, due to the use of considerably different sets of sequences and different methods of tree reconstruction employed, the results of different analyses are not directly comparable. The only observation that seems to be fully consistent is the grouping of eukaryotic polymerases δ and ζ . We made another attempt to reconstruct a phylogenetic tree of the family B DNAPs including only major branches of active polymerases (hence excluding PolB2) from archaea, eukaryotes and bacteria, and using an updated, representative set of sequences (Figure 4B). In the resulting tree, most of the deep branches are poorly resolved and unstable depending on the set of sequences used and the method of tree reconstruction (data not shown). The only additional observation that appears reliable is the confident grouping of PolB3 from several Methanomicrobia with the eukaryotic branch containing DNAPs δ and ζ (Figure 4B); this affinity is supported by the relatively high BLAST scores of the pairwise alignments of these sequences to eukaryotic polymerases compared with other archaeal polymerases (Supplementary Table S2). However, the PolB3 sequences from these Methanomicrobia lack the two Zn fingers at the C-terminus, a synapomorphy of the eukaryotic family B DNAPs that is also present in the archaeal PolD (Tahirov et al., 2009) (Figures 1, 4A and see discussion below). If grouping of PolB3 from Methanomicrobia with DNAPs δ and ζ reflects an actual evolutionary event, then a complicated scenario would have to be proposed, including acquisition of a eukaryotic polymerase by the ancestor of this group of organisms, loss of the “original” PolB3 and loss of the C-terminal Zn-fingers in the acquired polymerase. The alternative is the even less plausible scenario whereby the common ancestor of the eukaryotic DNAPs δ and ζ evolved from an unknown variant of the methanomicrobial PolB3 that contained at least one Zn finger; however, this scenario contradicts the recent conclusions on the origin of eukaryotes from the archaeal TACK superphylum (Martijn and Ettema, 2013; Koonin and Yutin, 2014). Given the complexity of these scenarios, the possibility should be considered that, the apparent strong support notwithstanding, the eukaryote-methanomicrobial affinity is yet another tree reconstruction artifact caused by large differences in evolutionary rates in different branches.

Thus, we have to conclude that phylogenetic analysis fails to resolve the evolutionary relationships between archaeal and eukaryotic family B DNAPs. So could any other considerations help understanding the origin of family B DNAPs that are responsible for eukaryotic DNA replication? In particular, this puzzle cannot be solved in full without uncovering the provenance of the family D polymerases, especially taking into account that the DP1 subunit clearly made it to LECA and is an indispensable component of all replicative B-family polymerases in eukaryotes (Yamasaki et al., 2010) whereas the DP2 subunit appears to have been lost. Furthermore, there is a significant, specific sequence similarity between the C-terminal Zn fingers of Pol- ϵ and DP2 (Tahirov et al., 2009). Any scenarios that strive to accommodate

all these findings require intricate chains of events (Tahirov et al., 2009).

An intriguing possibility is suggested by the conservation of several aspartate residues in the catalytic domain of DP2, including the DxD motif that is present in all palm domain polymerases and is involved in the binding of an essential divalent cation (Cann et al., 1998). This observation might indicate that, notwithstanding the absence of readily detectable sequence similarity, DP2 is a highly derived homolog of family B DNAPs. This hypothesis appears to be able to accommodate all available facts in the simplest possible fashion. The fact that the small subunit of the family D DNAP, DP1, is the readily detectable ortholog of the B subunit that is shared by all eukaryotic family B polymerases is also compatible with this scenario. It has been shown that the eukaryotic Pol- ϵ consists of an N-terminal DNAP domain in which all major catalytic motifs of family B are conserved and a C-terminal DNAP domain in which most of these motifs are disrupted suggestive of inactivation (Tahirov et al., 2009). The present hypothesis could account for the origin of Pol- ϵ as a fusion of an ancestral form of DP2 (before its accelerated evolution period) that would give rise to the active, N-terminal domain of Pol- ϵ , followed by an inactivated PolB2 domain inserting between the active N-terminal domain and the Zn finger (Figure 4). The N-terminal polymerase domain of Pol- ϵ shows a pattern of insertions and deletions that is distinct from those in all other family B DNAPs, which is compatible with a distinct origin (Tahirov et al., 2009). The accelerated evolution of the PolB1 at the origin of DP2 might have occurred within a viral genome, followed by reintroduction of this evolved gene into an ancestral euryarchaeal lineage via the so-called host-to-virus-to-host transfer loop, as has been proposed for the replicative MCM helicases of Methanococcales (Krupovic et al., 2010). In functional terms, this hypothesis is compatible with the fact that Pol- ϵ is the leading strand polymerase in eukaryotes. Obviously, this hypothesis will be put to test when the structure of the catalytic domain of DP2 is solved. Furthermore, the possibility remains that genome sequencing of currently uncharacterized, deep branches of archaea results in identification of novel DNAPs that help clarifying the relationships between the B and D families, and possibly, other aspects of DNAP evolution.

CONCLUSIONS

The DNAPs comprise the core of the DNA replication machinery, obviously one of the key functions in any cellular life form and many viruses and mobile elements. Other genes involved in key information processing functions, such as the core components of the translation and transcription systems, have highly conserved sequences, are rarely duplicated and do not seem to experience major accelerations of evolution (Koonin, 2003; Puigbo et al., 2009). Therefore, reconstruction of the evolution of the respective systems is a relatively straightforward task. Their major biological importance notwithstanding, the DNAPs evolve under a different regime that appears to involve multiple duplications, gene losses, horizontal gene transfers and domain rearrangements. Moreover, inactivated DNAPs seem to have adopted new functions the exact nature of which remains to be elucidated. The complexity of the evolution of the DNAPs is likely to stem partly from the functional

differentiation because in archaea and eukaryotes the lagging and leading strand are replicated by distinct DNAPs. Another important factor is the common presence of DNAPs in viruses and other mobile genetic elements that can transfer the DNAP genes between cellular organisms, providing an environment conducive to accelerated evolution, and possibly replacing the original genes.

The starker manifestation of the complexity of DNAP evolution is the intricate relationship between the archaeal and eukaryotic replication machineries. Here we proposed a parsimonious evolutionary scenario under which the archaeal family D of DNAPs is a highly derived form of family B. However, the available data are also compatible with various other scenarios that would involve contributions from different archaeal DNAPs and possibly also viruses.

ACKNOWLEDGMENTS

Eugene V. Koonin and Kira S. Makarova are supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine). Mart Krupovic was partly supported by the European Molecular Biology Organization (ASTF 82-2014).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00354/abstract>

REFERENCES

- Aravind, L., and Koonin, E. V. (1998). Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.* 26, 3746–3752. doi: 10.1093/nar/26.16.3746
- Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M. L. (2006). His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* 350, 228–239. doi: 10.1016/j.virol.2006.02.005
- Bauer, R. J., Begley, M. T., and Trakselis, M. A. (2012). Kinetics and fidelity of polymerization by DNA polymerase III from *Sulfolobus solfataricus*. *Biochemistry* 51, 1996–2007. doi: 10.1021/bi201799a
- Bebenek, K., Pedersen, L. C., and Kunkel, T. A. (2014). Structure-Function Studies of DNA Polymerase lambda. *Biochemistry* 53, 2781–2792. doi: 10.1021/bi4017236
- Bell, S. P., and Dutta, A. (2002). DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* 71, 333–374. doi: 10.1146/annurev.biochem.71.110601.135425
- Burgers, P. M., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., et al. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.* 276, 43487–43490. doi: 10.1074/jbc.R100056200
- Cann, I. K., Komori, K., Toh, H., Kanai, S., and Ishino, Y. (1998). A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14250–14255. doi: 10.1073/pnas.95.24.14250
- Choi, J. J., Nam, K. H., Min, B., Kim, S. J., Soll, D., and Kwon, S. T. (2006). Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. *J. Mol. Biol.* 356, 1093–1106. doi: 10.1016/j.jmb.2005.12.036
- Choi, J. Y., Eoff, R. L., Pence, M. G., Wang, J., Martin, M. V., Kim, E. J., et al. (2011). Roles of the four DNA polymerases of the crenarchaeon *Sulfolobus solfataricus* and accessory proteins in DNA replication. *J. Biol. Chem.* 286, 31180–31193. doi: 10.1074/jbc.M111.258038
- Cubanova, L., Richardson, T., Burkhardt, B. W., Kelman, Z., Connolly, B. A., Reeve, J. N., et al. (2013). Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *J. Bacteriol.* 195, 2322–2328. doi: 10.1128/JB.02037-12
- Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23, 324–328. doi: 10.1016/S0968-0004(98)01274-2
- Delagoutte, E. (2012). DNA polymerases: mechanistic insight from biochemical and biophysical studies. *Front. Biosci. (Landmark Ed.)* 17, 509–544. doi: 10.2741/3941
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgell, D. R., Malik, S. B., and Doolittle, W. F. (1998). Evidence of independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases. *Mol. Biol. Evol.* 15, 1207–1217. doi: 10.1093/oxfordjournals.molbev.a026028
- Filee, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002). Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *J. Mol. Evol.* 54, 763–773. doi: 10.1007/s00239-001-0078-x
- Forterre, P., and Prangishvili, D. (2013). The major role of viruses in cellular evolution: facts and hypotheses. *Curr. Opin. Virol.* 3, 558–565. doi: 10.1016/j.coviro.2013.06.013
- Gawel, D., Pham, P. T., Fijalkowska, I. J., Jonczyk, P., and Schaaper, R. M. (2008). Role of accessory DNA polymerases in DNA replication in *Escherichia coli*: analysis of the dnaX36 mutator mutant. *J. Bacteriol.* 190, 1730–1742. doi: 10.1128/JB.01463-07
- Gill, S., Krupovic, M., Desnoues, N., Beguin, P., Sezonov, G., and Forterre, P. (2014). A highly divergent archaeo-eukaryotic primase from the *Thermococcus nautilus* plasmid, pTN2. *Nucleic Acids Res.* 42, 3707–3719. doi: 10.1093/nar/gkt1385
- Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* 53, 263–287. doi: 10.1146/annurev.micro.56.012302.160741
- Guy, L., and Ettema, T. J. (2011). The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends Microbiol.* 19, 580–587. doi: 10.1016/j.tim.2011.09.002
- Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature* 503, 544–547. doi: 10.1038/nature12650
- Henneke, G., Flament, D., Hubscher, U., Querellou, J., and Raffin, J. P. (2005). The hyperthermophilic euryarchaeota Pyrococcus abyssi likely requires the two DNA polymerases D and B for DNA replication. *J. Mol. Biol.* 350, 53–64. doi: 10.1016/j.jmb.2005.04.042
- Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., et al. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3600–3605. doi: 10.1073/pnas.96.7.3600
- Iyer, L. M., Koonin, E. V., Leipe, D. D., and Aravind, L. (2005). Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.* 33, 3875–3896. doi: 10.1093/nar/gki702
- Jarosz, D. F., Beuning, P. J., Cohen, S. E., and Walker, G. C. (2007). Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol.* 15, 70–77. doi: 10.1016/j.tim.2006.12.004
- Jokela, M., Eskelinen, A., Pospiech, H., Rouvinen, J., and Syvoja, J. E. (2004). Characterization of the 3' exonuclease subunit DP1 of *Methanococcus jannaschii* replicative DNA polymerase D. *Nucleic Acids Res.* 32, 2430–2440. doi: 10.1093/nar/gkh558
- Kelman, Z., Pietrovsková, S., and Hurwitz, J. (1999). Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* deltaH. *J. Biol. Chem.* 274, 28751–28761. doi: 10.1074/jbc.274.40.28751
- Klinge, S., Nunez-Ramirez, R., Llorca, O., and Pellegrini, L. (2009). 3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases. *Embo J.* 28, 1978–1987. doi: 10.1038/embj.2009.150
- Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136. doi: 10.1038/nrmicro751
- Koonin, E. V., and Yutin, N. (2014). The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* 6:a016188. doi: 10.1101/cshperspect.a016188
- Krupovic, M., Gonnet, M., Hania, W. B., Forterre, P., and Erauso, G. (2013). Insights into dynamics of mobile genetic elements in hyperthermophilic

- environments from five new *Thermococcus* plasmids. *PLoS ONE* 8:e49044. doi: 10.1371/journal.pone.0049044
- Krupovic, M., Gribaldo, S., Bamford, D. H., and Forterre, P. (2010). The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol. Biol. Evol.* 27, 2716–2732. doi: 10.1093/molbev/msq161
- Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D., and Koonin, E. V. (2014a). Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* 12:36. doi: 10.1186/1741-7007-12-36
- Krupovic, M., Quemin, E. R., Bamford, D. H., Forterre, P., and Prangishvili, D. (2014b). Unification of the globally distributed spindle-shaped viruses of the Archaea. *J. Virol.* 88, 2354–2358. doi: 10.1128/JVI.02941-13
- Kunkel, T. A., and Burgers, P. M. (2008). Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol.* 18, 521–527. doi: 10.1016/j.tcb.2008.08.005
- Li, Z., Kelman, L. M., and Kelman, Z. (2013). *Thermococcus kodakarensis* DNA replication. *Biochem. Soc. Trans.* 41, 332–338. doi: 10.1042/BST20120303
- Lipps, G. (2011). Structure and function of the primase domain of the replication protein from the archaeal plasmid pRN1. *Biochem. Soc. Trans.* 39, 104–106. doi: 10.1042/BST0390104
- MacNeill, S. A. (2009). The haloarchaeal chromosome replication machinery. *Biochem. Soc. Trans.* 37, 108–113. doi: 10.1042/BST0370108
- Makarova, K. S., and Koonin, E. V. (2013). Archaeology of eukaryotic DNA replication. *Cold Spring Harb. Perspect. Biol.* 5:a012963. doi: 10.1101/cshperspect.a012963
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). The basic building blocks and evolution of CRISPR-Cas systems. *Biochem. Soc. Trans.* 41, 1392–1400. doi: 10.1042/BST20130038
- Martijn, J., and Ettema, T. J. (2013). From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* 41, 451–457. doi: 10.1042/BST20120292
- Matsui, I., Urushibata, Y., Shen, Y., Matsui, E., and Yokoyama, H. (2011). Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from *Pyrococcus horikoshii*. *FEBS Lett.* 585, 452–458. doi: 10.1016/j.febslet.2010.12.040
- Moon, A. F., Garcia-Diaz, M., Batra, V. K., Beard, W. A., Bebenek, K., Kunkel, T. A., et al. (2007). The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst)* 6, 1709–1725. doi: 10.1016/j.dnarep.2007.05.009
- Pavlov, Y. I., and Shcherbakova, P. V. (2010). DNA polymerases at the eukaryotic fork-20 years later. *Mutat. Res.* 685, 45–53. doi: 10.1016/j.mrfmmm.2009.08.002
- Peng, X., Basta, T., Haring, M., Garrett, R. A., and Prangishvili, D. (2007). Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. *Virology* 364, 237–243. doi: 10.1016/j.virol.2007.03.005
- Perler, F. B. (2002). InBase: the Intein Database. *Nucleic Acids Res.* 30, 383–384. doi: 10.1093/nar/30.1.383
- Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., et al. (1994). Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucleic Acids Res.* 22, 1125–1127. doi: 10.1093/nar/22.7.1125
- Pietila, M. K., Atanasova, N. S., Manole, V., Liljeroos, L., Butcher, S. J., Oksanen, H. M., et al. (2012). Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. *J. Virol.* 86, 5067–5079. doi: 10.1128/JVI.0915-11
- Pietila, M. K., Laurinmaki, P., Russell, D. A., Ko, C. C., Jacobs-Sera, D., Butcher, S. J., et al. (2013). Insights into head-tailed viruses infecting extremely halophilic archaea. *J. Virol.* 87, 3248–3260. doi: 10.1128/JVI.03397-12
- Podar, M., Makarova, K. S., Graham, D. E., Wolf, Y. I., Koonin, E. V., and Reysenbach, A. L. (2013). Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol. Direct* 8:9. doi: 10.1186/1745-6150-8-9
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490
- Puigbo, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a ‘Tree of Life’ in the thicket of the phylogenetic forest. *J. Biol.* 8, 59. doi: 10.1186/jbiol159
- Quax, T. E., Wolf, Y. I., Koehorst, J. J., Wurtzel, O., van der Oost, R., Ran, W., et al. (2013). Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep.* 4, 938–944. doi: 10.1016/j.celrep.2013.07.049
- Raymann, K., Forterre, P., Brochier-Armanet, C., and Gribaldo, S. (2014). Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol. Evol.* 6, 192–212. doi: 10.1093/gbe/evu004
- Rogozin, I. B., Makarova, K. S., Pavlov, Y. I., and Koonin, E. V. (2008). A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol. Direct* 3:32. doi: 10.1186/1745-6150-3-32
- Rothwell, P. J., and Waksman, G. (2005). Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* 71, 401–440. doi: 10.1016/S0065-3233(04)71011-6
- Sencilo, A., Jacobs-Sera, D., Russell, D. A., Ko, C. C., Bowman, C. A., Atanasova, N. S., et al. (2013). Snapshot of haloarchaeal tailed virus genomes. *RNA Biol.* 10, 803–816. doi: 10.4161/rna.24045
- Sharma, S., Helchowski, C. M., and Canman, C. E. (2013). The roles of DNA polymerase zeta and the Y family DNA polymerases in promoting or preventing genome instability. *Mutat. Res.* 743–744, 97–110. doi: 10.1016/j.mrfmmm.2012.11.002
- Shen, Y., Musti, K., Hiramoto, M., Kikuchi, H., Kawarabayashi, Y., and Matsui, I. (2001). Invariant Asp-1122 and Asp-1124 are essential residues for polymerization catalysis of family D DNA polymerase from *Pyrococcus horikoshii*. *J. Biol. Chem.* 276, 27376–27383. doi: 10.1074/jbc.M011762200
- Silverstein, T. D., Johnson, R. E., Jain, R., Prakash, L., Prakash, S., and Aggarwal, A. K. (2010). Structural basis for the suppression of skin cancers by DNA polymerase eta. *Nature* 465, 1039–1043. doi: 10.1038/nature09104
- Steitz, T. A., and Yin, Y. W. (2004). Accuracy, lesion bypass, strand displacement and translocation by DNA polymerases. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 359, 17–23. doi: 10.1098/rstb.2003.1374
- Tahirov, T. H., Makarova, K. S., Rogozin, I. B., Pavlov, Y. I., and Koonin, E. V. (2009). 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* 4:11. doi: 10.1186/1745-6150-4-11
- Tang, S. L., Nuttall, S., Ngui, K., Fisher, C., Lopez, P., and Dyall-Smith, M. (2002). HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. *Mol. Microbiol.* 44, 283–296. doi: 10.1046/j.1365-2958.2002.02890.x
- Wolf, Y. I., Makarova, K. S., Yutin, N., and Koonin, E. V. (2012). Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* 7:46. doi: 10.1186/1745-6150-7-46
- Yamasaki, K., Urushibata, Y., Yamasaki, T., Arisaka, F., and Matsui, I. (2010). Solution structure of the N-terminal domain of the archaeal D-family DNA polymerase small subunit reveals evolutionary relationship to eukaryotic B-family polymerases. *FEBS Lett.* 584, 3370–3375. doi: 10.1016/j.febslet.2010.06.026
- Yang, W. (2014). An Overview of Y-Family DNA Polymerases and a Case Study of Human DNA Polymerase eta. *Biochemistry* 53, 2793–2803. doi: 10.1021/bi500019s
- Yutin, N., Puigbo, P., Koonin, E. V., and Wolf, Y. I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE* 7:e36972. doi: 10.1371/journal.pone.0036972
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 29 May 2014; **accepted:** 24 June 2014; **published online:** 21 July 2014.
- Citation:** Makarova KS, Krupovic M and Koonin EV (2014) Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front. Microbiol.* 5:354. doi: 10.3389/fmicb.2014.00354
- This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.
- Copyright © 2014 Makarova, Krupovic and Koonin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Structural insights into eukaryotic DNA replication

Sylvie Doublie * and Karl E. Zahn

Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Erik Johansson, Umeå University, Sweden

Aneel Aggarwal, Mount Sinai School of Medicine, USA

***Correspondence:**

Sylvie Doublie, Department of Microbiology and Molecular Genetics, University of Vermont, 89 Beaumont Drive, Given E314A, Burlington, VT 05403, USA
e-mail: sdoublie@uvm.edu

Three DNA polymerases of the B family function at the replication fork in eukaryotic cells: DNA polymerases α , δ , and ϵ . DNA polymerase α , an heterotetramer composed of two primase subunits and two polymerase subunits, initiates replication. DNA polymerases δ and ϵ elongate the primers generated by pol α . The DNA polymerase from bacteriophage RB69 has served as a model for eukaryotic B family polymerases for some time. The recent crystal structures of pol δ , α , and ϵ revealed similarities but also a number of unexpected differences between the eukaryotic polymerases and their bacteriophage counterpart, and also among the three yeast polymerases. This review will focus on their shared structural elements as well as the features that are unique to each of these polymerases.

Keywords: DNA polymerase, B family, eukaryotic replication, fidelity, proofreading

INTRODUCTION

Replication in the nucleus of eukaryotic cells employs three DNA polymerases: polymerase α , δ , and ϵ (Hubscher et al., 2002; Pavlov et al., 2006b; Kunkel and Burgers, 2008; Loeb and Monnat, 2008; Burgers, 2009; Pavlov and Shcherbakova, 2010; Lange et al., 2011). DNA synthesis is directional and proceeds from 5' to 3', where nucleophilic attack on the α phosphate of a nucleotide by the 3'OH of a primer results in the incorporation of a nucleoside monophosphate and release of pyrophosphate (Steitz, 1999). All DNA polymerases require a primer and a free 3'OH to conduct DNA synthesis, and pol α is no exception. Pol α is a heterotetramer composed of two primase subunits and two polymerase subunits. The primase subunits initiate DNA replication by synthesizing short (7–12 ribonucleotides) RNA primers, which are then extended by polymerase α (Pellegrini, 2012). DNA polymerase δ and ϵ elongate the primers generated by pol α in an accurate and processive manner (Kunkel, 2004, 2011; Pellegrini, 2012). In yeast, DNA polymerase δ has been shown to be essential for DNA synthesis of the lagging strand whereas pol ϵ appears to mainly function at the leading strand (Pursell et al., 2007; Nick McElhinny et al., 2008; Kunkel, 2011; Georgescu et al., 2014). In contrast, in the mitochondria replication is the responsibility of one sole polymerase, DNA polymerase γ (Lee et al., 2009).

DNA polymerases are grouped into seven families (A, B, C, D, X, Y, and RT). In eukaryotes the three nuclear replicative DNA polymerases happen to belong to the B family (Burgers et al., 2001; Patel and Loeb, 2001). There are now crystals structures of all three replicative DNA polymerases from yeast, which allow for the first time a comparison of their shared structural elements as well as a study of their unique features (Swan et al., 2009; Perera et al., 2013; Hogg et al., 2014; Jain et al., 2014a). All three replicative DNA polymerases are multi-subunit enzymes (**Table 1**) (Johansson and MacNeill, 2010; Pavlov and Shcherbakova, 2010; Makarova et al., 2014). The main focus of this review is on their catalytic domain, or subunit A.

OVERALL STRUCTURE OF B FAMILY POLYMERASES

All DNA polymerases share a common polymerase fold, which has been compared to a human right hand, composed of three subdomains: fingers, palm, and thumb (Steitz, 1999; Patel and Loeb, 2001). The palm, a highly conserved fold composed of four antiparallel β strands and two helices, harbors two strictly conserved catalytic aspartates located in motif A, DXXLYPS and motif C, DTDS (Delarue et al., 1990; Braithwaite and Ito, 1993). This RRM-like fold is shared by a very large group of enzymes, including DNA and RNA polymerases, reverse transcriptases, CRISPR polymerase, and even reverse (3'-5') transferases such as Thg1 (Anantharaman et al., 2010; Hyde et al., 2010). In contrast, the thumb and fingers subdomains exhibit substantially more structural diversity (Steitz, 1999). The fingers undergo a conformational change upon binding DNA and the correct incoming nucleotide. This movement allows residues in the fingers subdomain to come in contact with the nucleotide in the nascent base pair. The thumb holds the DNA duplex during replication and plays a part in processivity (Doublie and Ellenberger, 1998; Doublie et al., 1999).

Eukaryotic DNA polymerases α , δ , and ϵ share homology with many archaeal, bacterial, bacteriophage, and viral polymerases (Delarue et al., 1990; Braithwaite and Ito, 1993; Franklin et al., 2001; Firbank et al., 2008; Wang and Yang, 2009). Koonin and collaborators contributed a detailed phylogenetic analysis of archaeal DNA polymerases and their relationship with eukaryotic polymerases in this issue of Frontiers in Microbiology dedicated to polymerases Makarova et al. (2014).

All B family polymerases are composed of five subdomains: the fingers, thumb, and palm (described above) constituting the core of the enzyme, as well as an exonuclease domain and an N-terminal domain (NTD) (Franklin et al., 2001; Xia and Konigsberg, 2014) (**Figure 1**; Table S1). The exonuclease domain carries a 3'-5' proofreading activity, which removes misincorporated nucleotides. The exonuclease active site is located 40–45 Å away from the polymerase active site. The NTD seems to be

Table 1 | Eukaryotic DNA polymerases are multi-subunit enzymes.

Polymerase	Species		Function
Polymerase α	<i>H. sapiens</i>	<i>S. cerevisiae</i>	
Catalytic or A-subunit	POLA1 (p180)	POL1	Catalytic subunit; polymerase activity; inactivated exonuclease
B-subunit	POLA2 (p70)	POL12	Regulatory subunit
Primase small subunit	PRIM1 (p49)	PRI1	Primase
Primase large subunit	PRIM2 (p58)	PRI2	Primase
Polymerase δ	<i>H. sapiens</i>	<i>S. cerevisiae</i>	
Catalytic or A-subunit	POLD1 (p125)	POL3	Catalytic subunit; has both polymerase and exonuclease activity
B-subunit	POLD2 (p50)	POL31	Accessory subunit
C-subunit	POLD3 (p66 or p68)	POL32	Accessory subunit
D-subunit	POLD4 (p12)	–	Accessory subunit
Polymerase ϵ	<i>H. sapiens</i>	<i>S. cerevisiae</i>	
Catalytic or A-subunit	POLE or POLE1	POL2	Catalytic subunit; has both polymerase and exonuclease activity
B-subunit	POLE2	DPB2	Accessory subunit
C-subunit	POLE3 (p17; CHRAC17)	DPB3	Accessory subunit
D-subunit	POLE4 (p12)	DPB4	Accessory subunit

devoid of catalytic activity. In pol δ the NTD comprises three motifs: one has a topology resembling an OB fold, a single-stranded DNA binding motif, and another bears an RNA-binding motif (RNA Recognition Motif or RRM) (Swan et al., 2009). In bacteriophage T4, mutations in the NTD decrease expression of the polymerase (Hughes et al., 1987). In RB69 and T4, the gp43 polymerase binds its own messenger RNA, presumably through the NTD and represses translation (Petrov et al., 2002), which does not seem to be the case for pol δ (Swan et al., 2009). New data indicate that the NTD plays a role in polymerase stability and fidelity through its interactions with other domains (Li et al., 2010; Prindle et al., 2013) (see below).

All mammalian B family DNA polymerases are known to harbor two cysteine-rich metal binding sites (CysA and CysB) in their C-terminal domain (CTD) (Figure 2). CysA is presumed to be a zinc-binding site whereas CysB is an iron sulfur cluster [4Fe-4S] (Netz et al., 2012). Loss of the [4Fe-4S] cluster in the CTD of yeast pol δ negatively affects interactions with its accessory B-subunit (Sanchez Garcia et al., 2004). The zinc-binding motif was shown to be important for interaction of pol δ with its processivity factor, PCNA (Netz et al., 2012).

DNA POLYMERASE α

The catalytic subunit of DNA polymerase α is composed of 1468 amino acids (Table 2). The protein construct designed for

crystallization was truncated at the N- and C-termini (residues 349–1258) and therefore lacks the CTD and its [4Fe-4S] cluster (Figure 2). The construct was crystallized unliganded, in a binary complex with a DNA/RNA hybrid oligonucleotide, and in a ternary complex with DNA/RNA and incoming nucleotide (Perera et al., 2013) (Figure 1).

A MECHANISM OF DISENGAGEMENT OF THE POLYMERASE

The RNA/DNA oligonucleotide captured in the crystals adopts an A-form conformation, as expected. The thumb domain engages in multiple interactions with the RNA primer, both via hydrophobic contacts and polar interactions (Perera et al., 2013). Experiments in solution have shown that the extension of the RNA primer by pol α is limited to 10–12 nucleotides, which amounts to one turn of a helix. This observation led the authors to suggest a mechanism for termination of primer synthesis by pol α in which loss of specific interactions between the thumb and the RNA would trigger the polymerase to disengage from the DNA/RNA oligonucleotide, and allow a hand off to a replicative polymerase.

MOVEMENTS IN THE PALM DOMAIN MAY FACILITATE TRANSLOCATION OF POL α

Having crystallized the enzyme in three states (apo, binary, and ternary) allowed the authors to overlay all three structural models. Pol α is the only eukaryotic family B DNA polymerase for which all three states were captured in a crystal structure. The structural superposition revealed that, in addition to the well-documented movements of the fingers and thumb subdomains accompanying substrate binding and nucleotidyl transfer, the palm subdomain itself undergoes a structural rearrangement (Perera et al., 2013). The authors propose that the different conformations of the palm domain could facilitate translocation of pol α along and beyond the RNA/DNA duplex. As mentioned above, loss of contacts to the RNA strand is predicted to trigger release of primer, which then becomes available for extension by pol δ or ϵ .

A DIFFERENT PROTEIN FOLD IN THE INACTIVATED EXONUCLEASE SUBDOMAIN

The proofreading activity is abolished in pol α , due to mutations in all four carboxylates (Asp114/Glu116/Asp222/Asp327 in RB69 gp43 correspond to Ser542/Gln544/Tyr644/Asn757 in a structure-based alignment) (Table 2). Moreover, the β -hairpin motif found in most polymerases of the B family (residues 246–267 in RB69 gp43) is replaced by a helical region in pol α (residues 667–676; 681–693) (Hogg et al., 2007). The β hairpin is part of the exonuclease domain and has been shown in T4 and RB69 pols to participate in the partitioning of the DNA primer between the polymerase and the exonuclease active site (Reha-Krantz, 1988; Stocki et al., 1995; Hogg et al., 2007). In the absence of proofreading activity it is not surprising that this motif was not retained in pol α . Residues His 684 and Phe 685 of the helical region in pol α stack with a thymine and guanine base, respectively, at positions -3 and -2 in the unpaired 5'end of the template (Perera et al., 2013). Thus, in pol α the region corresponding to the β -hairpin motif adopts a different fold (helices vs. β strands) and a different function (stabilizing the unpaired region of the template strand rather than facilitating active site

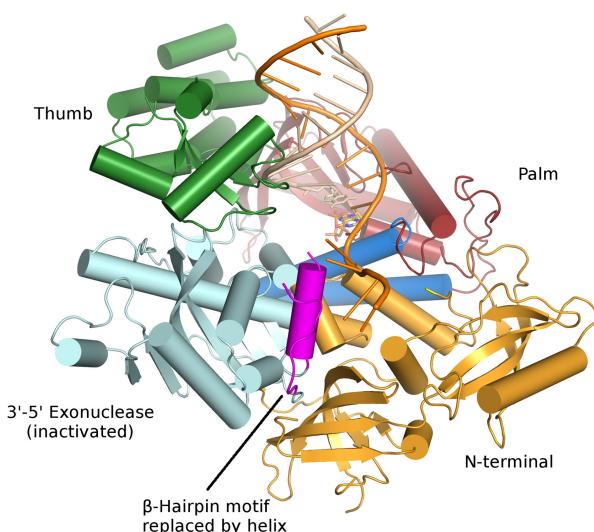
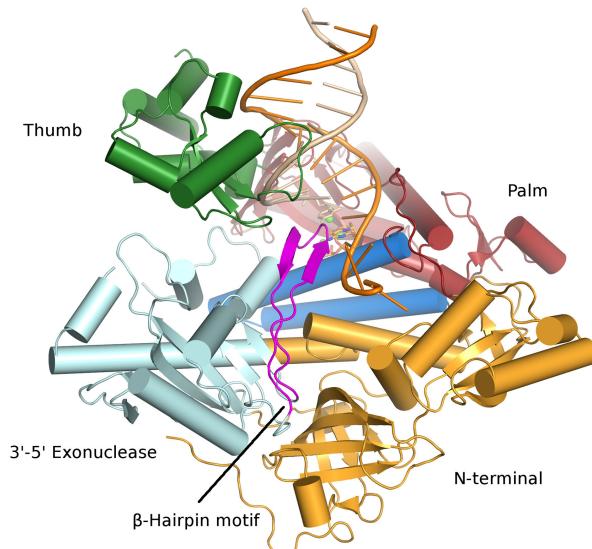
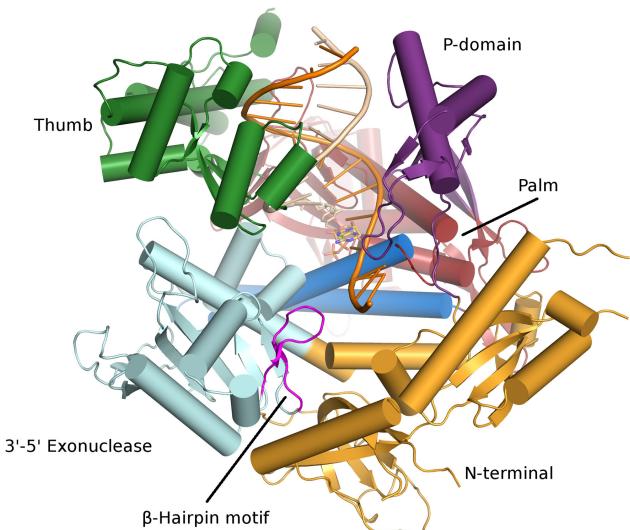
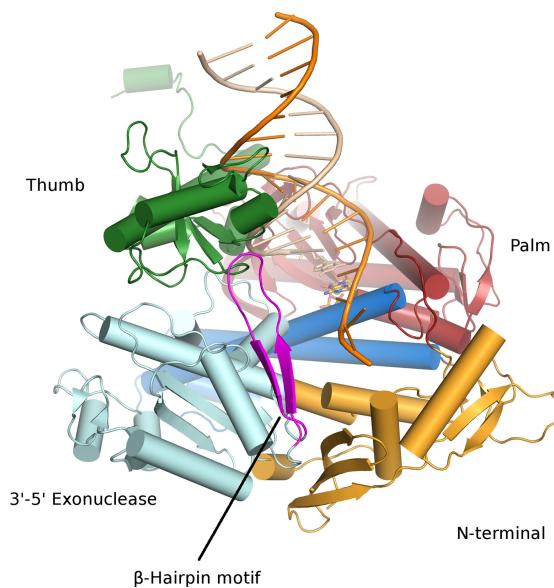
A Polymerase α **B Polymerase δ** **C Polymerase ϵ** **D RB69 gp43 polymerase**

FIGURE 1 | Ternary complexes of polymerases α , δ , ϵ , and RB69 gp43 are illustrated from identical orientations for comparison. The thumb (green) and fingers (dark blue) domains grasp the duplex nucleic acid (primer shown in beige, template in orange) against the palm domain (red). The N-terminal domain appears in gold, adjacent to the 3'-5' exonuclease domain (cyan). **(A)** Polymerase α (PDBID 4FYD) binds an RNA/DNA hybrid, where the wide, shallow minor groove of A-form DNA is apparent near the thumb. The 3'-5' exonuclease domain is devoid of activity. A helical region (magenta) in the inactivated exonuclease domain stabilizes the 5'end of the template. **(B)** Polymerase δ (PDBID 3IAY) harbors a large β hairpin motif (magenta), which is

important in switching the primer strand from the polymerase active site to the exonuclease active site in the event of proofreading. **(C)** Polymerase ϵ (PDBID 4M8O) yields a unique P-domain (purple), which endows the polymerase with increased processivity. Interestingly, the β hairpin motif is atrophied in pol ϵ . **(D)** Conservation of the family B DNA polymerase fold, and domain organization, is evident when the model enzyme from bacteriophage RB69 gp43 (PDBID 2OZS) is viewed along with the three eukaryotic replicative polymerases. The domain delineation for each polymerase is given in Table S1. Figure was made with PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.).

switching). Since pol α is devoid of proofreading activity the question arises as to whether the short oligonucleotides are corrected, and if so, by which DNA polymerase. It appears that proofreading of the primers synthesized by pol α is performed by pol δ (Pavlov et al., 2006a).

DNA POLYMERASE δ

Human pol δ is composed of four subunits whereas *Saccharomyces cerevisiae* has three (Geric et al., 1998; Liu et al., 2000) (Table 1). In addition to its function in DNA replication pol δ has been shown to play a role in DNA repair and recombination (Hubscher

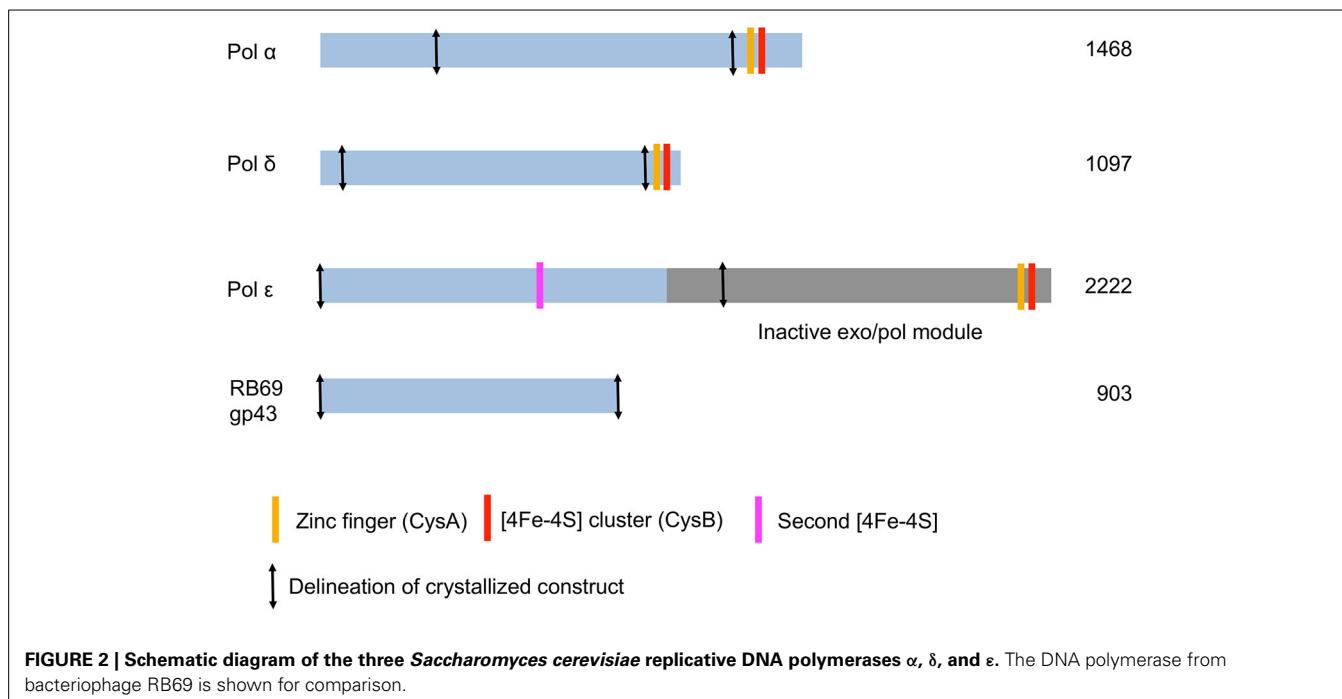


FIGURE 2 | Schematic diagram of the three *Saccharomyces cerevisiae* replicative DNA polymerases α , δ , and ϵ . The DNA polymerase from bacteriophage RB69 is shown for comparison.

et al., 2002; Lee et al., 2012; Tahirov, 2012). P12, the smallest subunit in human pol δ and also the subunit that is not seen in budding yeast, is degraded in response to DNA damage (Lee et al., 2014). The catalytic subunit of yeast pol δ (POL3) is composed of 1097 residues. The construct used for crystallization comprises residues 67–985 and thus lacks the CTD (Figure 1; Table 2).

A THIRD METAL ION IN THE POLYMERASE ACTIVE SITE

The palm domain contains three conserved carboxylates (Asp608, Asp762, and Asp764). The two catalytic aspartates, Asp608 and Asp764, contact two metal ions (Ca^{2+}) in the polymerase active site separated by 3.7 Å. Intriguingly a third metal was observed coordinated by the γ phosphate of the incoming nucleotide and Glu802, with Glu800 in the vicinity. Mutating both glutamates to alanine yielded a polymerase variant with reduced incorporation efficiency for both correct and incorrect nucleotides (Swan et al., 2009). At these amino acid positions, pol α and pol ϵ also have carboxylate residues (pol δ Glu800/Glu802 correspond to pol α Asp1033/Asp1035, and pol ϵ Glu945/Asp947). Whether these carboxylates play similar roles in pol α and ϵ remains to be investigated.

HIGH FIDELITY AND PROOFREADING

Human pol δ is a high-fidelity polymerase, catalyzing the nucleotidyl transfer reaction with an error frequency of 1 per 22,000 (Schmitt et al., 2009). Proofreading boosts the fidelity of the polymerase by a factor of 10–100 (McCulloch and Kunkel, 2008; Prindle et al., 2013). Pol δ harbors a polymerase and exonuclease active site, separated by about 45 Å (Swan et al., 2009). DNA polymerases with proofreading activity are able to sense misincorporated nucleotides by contacting the minor groove of base pairs beyond the insertion site. The protein interacts with universal

hydrogen bond acceptors at the N3 and O2 positions of purines and pyrimidines, respectively (Seeman et al., 1976; Doublie et al., 1998; Franklin et al., 2001). These hydrogen bond contacts are preserved when the base pair adopts a Watson-Crick geometry and lost in the event of a mismatch. In RB69 gp43, the contacts extend to the first two base pairs beyond the nascent base pair (Franklin et al., 2001; Hogg et al., 2004, 2005). The contacts are much more extensive in pol δ , extending to five base pairs post-insertion (Swan et al., 2009), which could contribute to its high fidelity.

As mentioned above, the β -hairpin segment from the exonuclease domain plays a critical role in the partition of the DNA between polymerization and proofreading sites in T4 and RB69 pols (Stocki et al., 1995; Hogg et al., 2007). In RB69 gp43 the β -hairpin motif adopts different conformations, depending on whether the complex was obtained with undamaged DNA (Franklin et al., 2001; Zahn et al., 2007) or DNA containing a damage (Freisinger et al., 2004; Hogg et al., 2004). It was fully visualized contacting both the primer and template strands in a complex with thymine glycol (Aller et al., 2011). Similarly, the β hairpin in pol δ protrudes into the major groove of the DNA and acts as a wedge between double-stranded DNA and the single-stranded 5' end of the template strand, which is stabilized by two aromatic residues Phe441 and Tyr446 (Figure 1) (Swan et al., 2009). The position of the β hairpin is consistent with a role in active site switching.

INTERDOMAIN CONTACTS AND FIDELITY

Mutations involved in cancer are mostly found in the exonuclease domain of pol δ and ϵ , emphasizing the critical role of proofreading in lowering the incidence of mutations (Church et al., 2013; Henninger and Pursell, 2014). One mutation in human colorectal cancer cells localizes to the fingers domain, R689W. The

Table 2 | Compilation of DNA polymerases of the B family of known structure.

Polymerase name	Yeast gene	Conserved aspartates in polymerase active site	Conserved carboxylates in exonuclease active site	UNIPROT ID	Source	Iron sulfur clusters?	Number of amino acids	Protein construct crystallized	PDB ID code and reference
Pol α	POL1	Asp864 Asp998 Asp996	N/A	P13382	<i>S. cerevisiae</i>	1 [4Fe-4S] Cys1348,1353,1367,1372	1468 aa	349–1258	4FYD 3.1 Å Perera et al., 2013
Pol δ	POL3	Asp608 Asp764 Asp762	Asp321 Glu323 Asp407 Asp520	P15436	<i>S. cerevisiae</i>	1 [4Fe-4S] Cys1056,1059,1069,1074	1097 aa	67–985	3IAY 2.0 Å Swan et al., 2009
Pol ϵ	POL2	Asp640 Asp877 Asp875	Asp290 Glu292 Asp383 Asp477	P21951	<i>S. cerevisiae</i>	2 [4Fe-4S] Cys665,667,668,763 Cys2164,2167,2179,2181	2222 aa	1–1228 1–1187	4M8O 2.2 Å (Hogg et al., 2014) 4PTF 2.8 Å Jain et al., 2014a
RB69 gp43		Asp411 Asp623 Asp621	Asp114 Glu116 Asp222 Asp327	Q38087	Bacteriophage RB69	None	903 aa	1–903	1IG9 2.6 Å Franklin et al., 2001
<i>E. coli</i> Pol II		Asp419 Asp547 Asp545	Asp156 Glu158 Asp229 Asp315	P21189	<i>Escherichia coli</i>	None	783 aa	1–780	3K57 2.08 Å Wang and Yang, 2009
Archaeal Tgo		Asp404 Asp542 Asp540	Asp141 Glu143 Asp215 Asp335	P56689	<i>Thermococcus gorgonarius</i>	None	773 aa	1–757	2VWJ 2.78 Å Firbank et al., 2008

analogous mutation in yeast (R696W) results in a mutator phenotype (Daee et al., 2010). A mutation in the vicinity of Arg696 in the highly conserved motif B of the fingers subdomain of yeast pol δ (A699Q) also results in a mutator phenotype. This region of the fingers is in close proximity to the NTD. Mutating Met540 of the NTD to alanine abolishes the mutator phenotype of A699Q, illustrating that interactions between the fingers and the NTD can affect the fidelity of the polymerase (Prindle et al., 2013). Similarly in T4 and RB69 pols the NPL core motif, which involves residues from the N-terminal and palm domains, is in contact with the fingers domain and was shown to stabilize polymerase-DNA complexes (Li et al., 2010).

DNA POLYMERASE ϵ

The catalytic subunit of DNA polymerase ϵ is the product of a very large gene (2222 amino acids in yeast; 2286 in humans), and is only third in size after polymerase ζ (also a member of the B family) and pol θ , a family A polymerase (3130 and 2590 amino acids, respectively, in humans) (Lange et al., 2011; Hogg and Johansson, 2012) (Figure 1; Table 2). Pol ϵ is twice as large as pol δ and is composed of two tandem polymerase/exonuclease regions. The N-terminal segment harbors both polymerase and proofreading activities whereas the C-terminal segment is inactivated. The two exonuclease-polymerase modules are distantly related (Tahirov et al., 2009). Although the inactivated segment is presumed to play a structural role during replication, two groups were able to crystallize catalytically active pol ϵ constructs (residues 1–1228; 1–1187) lacking the entire C-terminal module (Hogg et al., 2014; Jain et al., 2014a). Both crystal structures were of a ternary complex of the polymerase, DNA primer/template and incoming nucleotide.

A NOVEL PROCESSIVITY DOMAIN EMANATING FROM THE PALM DOMAIN

Pol ϵ differs from pol δ in that it does not require the DNA sliding clamp PCNA for high processivity (Hogg and Johansson, 2012). The palm domain of pol ϵ is substantially larger (380 residues) than that of pol α or δ (175 and 203 residues, respectively). The recent pol ϵ crystal structures revealed that insertions in the palm domain collectively form a new domain consisting of three β strands and two helices (residues 533–555; 682–760) (Hogg et al., 2014; Jain et al., 2014a) (Figure 1; Table S1). Deleting residues 690–751 resulted in a variant with decreased polymerase activity. Moreover, mutating positively charged residues (His748, Arg749, and Lys751) located in the vicinity of the phosphate backbone affected the processivity of the enzyme (Hogg et al., 2014). The extra domain originating from the palm was thus named the processivity or P domain, after its function. The base of the P domain harbors a metal binding site (see below) (Hogg et al., 2014; Jain et al., 2014a,b).

AN IRON SULFUR CLUSTER WITHIN THE POLYMERASE DOMAIN

Unexpectedly solution studies revealed that the catalytic subunit of yeast polymerase ϵ itself contains an [4Fe-4S] cluster within its polymerase fold (Jain et al., 2014b), in addition to the [4Fe-4S] cluster in the CTD (Figure 2; Table 2). The second [4Fe-4S] cluster within pol ϵ suggests that this polymerase may be more

sensitive to oxidative stress (Jain et al., 2014b). The crystal structures of pol ϵ , however, did not reveal a [4Fe-4S] cluster in the polymerase domain (Hogg et al., 2014; Jain et al., 2014a; Zahn and Doublie, 2014). Two of the cysteines residues are disordered in the structural models and the resulting metal binding site appears to bind zinc (Hogg et al., 2014; Jain et al., 2014a). Substitution of a [4Fe-4S] by a non-native zinc in metal-binding proteins is not unusual (Netz et al., 2012) as [4Fe-4S] clusters are labile. Visualizing the [4Fe-4S] within the polymerase domain of pol ϵ may necessitate anaerobic conditions.

A SHORT β -HAIRPIN MOTIF IN THE EXONUCLEASE DOMAIN

In any DNA polymerase harboring both polymerase and exonuclease activities the bound DNA is in equilibrium between the two active centers (Beechem et al., 1998). The concentration of incoming nucleotide and the presence of a damaged base or mispair are two factors that influence the transfer of DNA from the polymerase active site to the proofreading active site. Polymerases monitor the minor groove side of the newly formed base pairs and interact with the universal H bond acceptors, O3, and N2, as a way of checking for mismatches (Seeman et al., 1976; Franklin et al., 2001). A unique feature of pol ϵ is the contact to the major groove side of the nascent base pair via a residue from the exonuclease domain, Tyr431. Further analysis is warranted to elucidate the potential role of this tyrosine in the high fidelity of pol ϵ .

In pol δ the β -hairpin segment inserts itself in the DNA and acts as a wedge between single-stranded and double-stranded DNA (Swan et al., 2009). In *E. coli* DNA pol II, the insertion of a β barrel shifts the position of the β hairpin in such a way that polymerization is favored over proofreading (Wang and Yang, 2009). This modification presumably allows this polymerase to carry out translesion synthesis extension. Since pol ϵ is an accurate DNA polymerase the assumption before knowledge of the crystal structure would be that the β hairpin should be closer to that of pol δ than that of *E. coli* Pol II. Surprisingly, the β -hairpin motif in pol ϵ is truncated, too short to contact the DNA (Figure 1). Which protein motif, then, might be facilitating active site switching upon sensing of a mispair? The P domain is a good candidate, because of its contacts to both primer and template strands; residues from the P domain could sense replication errors and thus may help facilitate active site switching.

CONCLUSIONS

All three eukaryotic replicative DNA polymerases use a common B-family fold, and each polymerase has incorporated modified structural elements which are unique and tailored for each polymerase's specific function (for example, the addition of the processivity domain in pol ϵ , a processive polymerase that does not use PCNA, or the modified region contacting the 5'end of the template in pol α , a polymerase devoid of proofreading activity). The fold of B family polymerases is well suited for high-fidelity, replicative polymerases. But surprisingly, it is also used by translesion polymerases. Eukaryotic pol ζ (or REV3L) is a 353 kDa polymerase which functions in translesion synthesis and appears to suppress tumorigenesis (Wittschieben et al., 2010; Lange et al., 2011; Zahn et al., 2011; Hogg and Johansson, 2012; Sharma et al.,

2013). The structure of *E. coli* Pol II revealed modifications in the NTD which affect the position of the β hairpin of the exonuclease domain, and thus partitioning of the DNA between the polymerization and proofreading sites (Wang and Yang, 2009). The structure of pol ζ may reveal similar adjustments, which alter the fold employed by high-fidelity, replicative polymerases to render the enzyme less faithful and able to perform translesion synthesis.

ACKNOWLEDGMENT

This work was supported by a grant from the National Institutes of Health (NCI R01 CA 52040).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/Journal/10.3389/fmicb.2014.00444/abstract>

REFERENCES

- Aller, P., Duclos, S., Wallace, S. S., and Doublie, S. (2011). A crystallographic study of the role of sequence context in thymine glycol bypass by a replicative DNA polymerase serendipitously sheds light on the exonuclease complex. *J. Mol. Biol.* 412, 22–34. doi: 10.1016/j.jmb.2011.07.007
- Anantharaman, V., Iyer, L. M., and Aravind, L. (2010). Presence of a classical RRM-fold palm domain in Thg1-type 3'-5' nucleic acid polymerases and the origin of the GGDEF and CRISPR polymerase domains. *Biol. Direct* 5, 43. doi: 10.1186/1745-6150-5-43
- Beechem, J. M., Otto, M. R., Bloom, L. B., Eritja, R., Reha-Krantz, L. J., and Goodman, M. F. (1998). Exonuclease-polymerase active site partitioning of primer-template DNA strands and equilibrium Mg²⁺ binding properties of bacteriophage T4 DNA polymerase. *Biochemistry* 37, 10144–10155. doi: 10.1021/bi980074b
- Braithwaite, D. K., and Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21, 787–802. doi: 10.1093/nar/21.4.787
- Burgers, P. M. (2009). Polymerase dynamics at the eukaryotic DNA replication fork. *J. Biol. Chem.* 284, 4041–4045. doi: 10.1074/jbc.R800062200
- Burgers, P. M., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., et al. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.* 276, 43487–43490. doi: 10.1074/jbc.R100056200
- Church, D. N., Briggs, S. E., Palles, C., Domingo, E., Kearsey, S. J., Grimes, J. M., et al. (2013). DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum. Mol. Genet.* 22, 2820–2828. doi: 10.1093/hmg/ddt131
- Daei, D. L., Mertz, T. M., and Shcherbakova, P. V. (2010). A cancer-associated DNA polymerase delta variant modeled in yeast causes a catastrophic increase in genomic instability. *Proc. Natl. Acad. Sci. U.S.A.* 107, 157–162. doi: 10.1073/pnas.0907526106
- Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* 3, 461–467. doi: 10.1093/protein/3.6.461
- Doublie, S., and Ellenberger, T. (1998). The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.* 8, 704–712. doi: 10.1016/S0959-440X(98)80089-4
- Doublie, S., Sawaya, M. R., and Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure* 7, R31–R35. doi: 10.1016/S0969-2126(99)80017-3
- Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391, 251–258. doi: 10.1038/34593
- Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J., and Connolly, B. A. (2008). Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.* 381, 529–539. doi: 10.1016/j.jmb.2008.06.004
- Franklin, M. C., Wang, J., and Steitz, T. A. (2001). Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* 105, 657–667. doi: 10.1016/S0092-8674(01)00367-1
- Freisinger, E., Grollman, A. P., Miller, H., and Kisker, C. (2004). Lesion (in)tolerance reveals insights into DNA replication fidelity. *EMBO J.* 23, 1494–1505. doi: 10.1038/sj.emboj.7600158
- Georgescu, R. E., Langston, L., Yao, N. Y., Yurieva, O., Zhang, D., Finkelstein, J., et al. (2014). Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* 21, 664–670. doi: 10.1038/nsmb.2851
- Gerik, K. J., Li, X., Pautz, A., and Burgers, P. M. (1998). Characterization of the two small subunits of *Saccharomyces cerevisiae* DNA polymerase delta. *J. Biol. Chem.* 273, 19747–19755. doi: 10.1074/jbc.273.31.19747
- Henninger, E. E., and Pursell, Z. F. (2014). DNA polymerase epsilon and its roles in genome stability. *IUBMB Life* 66, 339–351. doi: 10.1002/iub.1276
- Hogg, M., Aller, P., Konigsberg, W., Wallace, S. S., and Doublie, S. (2007). Structural and biochemical investigation of the role in proofreading of a beta hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. *J. Biol. Chem.* 282, 1432–1444. doi: 10.1074/jbc.M605675200
- Hogg, M., and Johansson, E. (2012). DNA polymerase epsilon. *Subcell Biochem.* 62, 237–257. doi: 10.1007/978-94-007-4572-8_13
- Hogg, M., Osterman, P., Bylund, G. O., Ganai, R. A., Lundstrom, E. B., Sauer-Eriksson, A. E., et al. (2014). Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. *Nat. Struct. Mol. Biol.* 21, 49–55. doi: 10.1038/nsmb.2712
- Hogg, M., Wallace, S. S., and Doublie, S. (2004). Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J.* 23, 1483–1493. doi: 10.1038/sj.emboj.7600150
- Hogg, M., Wallace, S. S., and Doublie, S. (2005). Bumps in the road: how replicative DNA polymerases see DNA damage. *Curr. Opin. Struct. Biol.* 15, 86–93. doi: 10.1016/j.sbi.2005.01.014
- Hubscher, U., Maga, G., and Spadari, S. (2002). Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 71, 133–163. doi: 10.1146/annurev.biochem.71.090501.150041
- Hughes, M. B., Yee, A. M., Dawson, M., and Karam, J. (1987). Genetic mapping of the amino-terminal domain of bacteriophage T4 DNA polymerase. *Genetics* 115, 393–403.
- Hyde, S. J., Eckenroth, B. E., Smith, B. A., Eberley, W. A., Heintz, N. H., Jackman, J. E., et al. (2010). tRNA(His) guanylyltransferase (THG1), a unique 3'-5' nucleotidyl transferase, shares unexpected structural homology with canonical 5'-3' DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20305–20310. doi: 10.1073/pnas.1010436107
- Jain, R., Rajashankar, K. R., Buku, A., Johnson, R. E., Prakash, L., Prakash, S., et al. (2014a). Crystal structure of yeast DNA polymerase epsilon catalytic domain. *PLoS ONE* 9:e94835. doi: 10.1371/journal.pone.0094835
- Jain, R., Vanamee, E. S., Dzikovski, B. G., Buku, A., Johnson, R. E., Prakash, L., et al. (2014b). An iron-sulfur cluster in the polymerase domain of yeast DNA polymerase epsilon. *J. Mol. Biol.* 426, 301–308. doi: 10.1016/j.jmb.2013.10.015
- Johansson, E., and Macneill, S. A. (2010). The eukaryotic replicative DNA polymerases take shape. *Trends Biochem. Sci.* 35, 339–347. doi: 10.1016/j.tibs.2010.01.004
- Kunkel, T. A. (2004). DNA replication fidelity. *J. Biol. Chem.* 279, 16895–16898. doi: 10.1074/jbc.R400006200
- Kunkel, T. A. (2011). Balancing eukaryotic replication asymmetry with replication fidelity. *Curr. Opin. Chem. Biol.* 15, 620–626. doi: 10.1016/j.cbpa.2011.07.025
- Kunkel, T. A., and Burgers, P. M. (2008). Dividing the workload at a eukaryotic replication fork. *Trends Cell. Biol.* 18, 521–527. doi: 10.1016/j.tcb.2008.08.005
- Lange, S. S., Takata, K., and Wood, R. D. (2011). DNA polymerases and cancer. *Nat. Rev. Cancer* 11, 96–110. doi: 10.1038/nrc2998
- Lee, M. Y., Zhang, S., Lin, S. H., Chea, J., Wang, X., Leroy, C., et al. (2012). Regulation of human DNA polymerase delta in the cellular responses to DNA damage. *Environ. Mol. Mutagen.* 53, 683–698. doi: 10.1002/em.21743
- Lee, M. Y., Zhang, S., Lin, S. H., Wang, X., Darzynkiewicz, Z., Zhang, Z., et al. (2014). The tail that wags the dog: p12, the smallest subunit of DNA polymerase delta, is degraded by ubiquitin ligases in response to DNA damage and during cell cycle progression. *Cell Cycle* 13, 23–31. doi: 10.4161/cc.27407
- Lee, Y. S., Kennedy, W. D., and Yin, Y. W. (2009). Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* 139, 312–324. doi: 10.1016/j.cell.2009.07.050
- Li, V., Hogg, M., and Reha-Krantz, L. J. (2010). Identification of a new motif in family B DNA polymerases by mutational analyses of the bacteriophage t4 DNA polymerase. *J. Mol. Biol.* 400, 295–308. doi: 10.1016/j.jmb.2010.05.030

- Liu, L., Mo, J., Rodriguez-Belmonte, E. M., and Lee, M. Y. (2000). Identification of a fourth subunit of mammalian DNA polymerase delta. *J. Biol. Chem.* 275, 18739–18744. doi: 10.1074/jbc.M001217200
- Loeb, L. A., and Monnat, R. J. Jr. (2008). DNA polymerases and human disease. *Nat. Rev. Genet.* 9, 594–604. doi: 10.1038/nrg2345
- Makarova, K. S., Krupovic, M., and Koonin, E. V. (2014). Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front Microbiol.* 5:354. doi: 10.3389/fmicb.2014.00354
- McCulloch, S. D., and Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* 18, 148–161. doi: 10.1038/cr.2008.4
- Netz, D. J., Stith, C. M., Stumpfig, M., Kopf, G., Vogel, D., Genau, H. M., et al. (2012). Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. *Nat. Chem. Biol.* 8, 125–132. doi: 10.1038/nchembio.721
- Nick Mcelhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M., and Kunkel, T. A. (2008). Division of labor at the eukaryotic replication fork. *Mol. Cell* 30, 137–144. doi: 10.1016/j.molcel.2008.02.022
- Patel, P. H., and Loeb, L. A. (2001). Getting a grip on how DNA polymerases function. *Nat. Struct. Biol.* 8, 656–659. doi: 10.1038/90344
- Pavlov, Y. I., Frahm, C., Nick Mcelhinny, S. A., Niimi, A., Suzuki, M., and Kunkel, T. A. (2006a). Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta. *Curr. Biol.* 16, 202–207. doi: 10.1016/j.cub.2005.12.002
- Pavlov, Y. I., Shcherbakova, P. V., and Rogozin, I. B. (2006b). Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. *Int. Rev. Cytol.* 255, 41–132. doi: 10.1016/S0074-7696(06)55002-8
- Pavlov, Y. I., and Shcherbakova, P. V. (2010). DNA polymerases at the eukaryotic fork-20 years later. *Mutat. Res.* 685, 45–53. doi: 10.1016/j.mrfmmm.2009.08.002
- Pellegrini, L. (2012). The Pol alpha-primase complex. *Subcell Biochem.* 62, 157–169. doi: 10.1007/978-94-007-4572-8_9
- Perera, R. L., Torella, R., Klinge, S., Kilkenny, M. L., Maman, J. D., and Pellegrini, L. (2013). Mechanism for priming DNA synthesis by yeast DNA Polymerase alpha. *eLife* 2, e00482. doi: 10.7554/eLife.00482
- Petrov, V. M., Ng, S. S., and Karam, J. D. (2002). Protein determinants of RNA binding by DNA polymerase of the T4-related bacteriophage RB69. *J. Biol. Chem.* 277, 33041–33048. doi: 10.1074/jbc.M204754200
- Prindle, M. J., Schmitt, M. W., Parmeggiani, F., and Loeb, L. A. (2013). A substitution in the fingers domain of DNA polymerase delta reduces fidelity by altering nucleotide discrimination in the catalytic site. *J. Biol. Chem.* 288, 5572–5580. doi: 10.1074/jbc.M112.436410
- Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E., and Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 317, 127–130. doi: 10.1126/science.1144067
- Reha-Krantz, L. J. (1988). Amino acid changes coded by bacteriophage T4 DNA polymerase mutator mutants. Relating structure to function. *J. Mol. Biol.* 202, 711–724. doi: 10.1016/0022-2836(88)90552-9
- Sanchez Garcia, J., Ciuffo, L. F., Yang, X., Kearsey, S. E., and Macneill, S. A. (2004). The C-terminal zinc finger of the catalytic subunit of DNA polymerase delta is responsible for direct interaction with the B-subunit. *Nucleic Acids Res.* 32, 3005–3016. doi: 10.1093/nar/gkh623
- Schmitt, M. W., Matsumoto, Y., and Loeb, L. A. (2009). High fidelity and lesion bypass capability of human DNA polymerase delta. *Biochimie* 91, 1163–1172. doi: 10.1016/j.biochi.2009.06.007
- Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U.S.A.* 73, 804–808. doi: 10.1073/pnas.73.3.804
- Sharma, S., Helchowski, C. M., and Canman, C. E. (2013). The roles of DNA polymerase zeta and the Y family DNA polymerases in promoting or preventing genome instability. *Mutat. Res.* 743–744, 97–110. doi: 10.1016/j.mrfmmm.2012.11.002
- Steitz, T. A. (1999). DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* 274, 17395–17398. doi: 10.1074/jbc.274.25.17395
- Stocki, S. A., Nonay, R. L., and Reha-Krantz, L. J. (1995). Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 3' → 5' exonuclease activities. *J. Mol. Biol.* 254, 15–28. doi: 10.1006/jmbi.1995.0595
- Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009). Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. *Nat. Struct. Mol. Biol.* 16, 979–986. doi: 10.1038/nsmb.1663
- Tahirov, T. H. (2012). Structure and function of eukaryotic DNA polymerase delta. *Subcell Biochem.* 62, 217–236. doi: 10.1007/978-94-007-4572-8_12
- Tahirov, T. H., Makarova, K. S., Rogozin, I. B., Pavlov, Y. I., and Koonin, E. V. (2009). 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* 4, 11. doi: 10.1186/1745-6150-4-11
- Wang, F., and Yang, W. (2009). Structural insight into translesion synthesis by DNA Pol II. *Cell* 139, 1279–1289. doi: 10.1016/j.cell.2009.11.043
- Wittschieben, J. P., Patil, V., Glushets, V., Robinson, L. J., Kusewitt, D. F., and Wood, R. D. (2010). Loss of DNA polymerase zeta enhances spontaneous tumorigenesis. *Cancer Res.* 70, 2770–2778. doi: 10.1158/0008-5472.CAN-09-4267
- Xia, S., and Konigsberg, W. H. (2014). RB69 DNA polymerase structure, kinetics, and fidelity. *Biochemistry* 53, 2752–2767. doi: 10.1021/bi4014215
- Zahn, K. E., Belrhali, H., Wallace, S. S., and Doublie, S. (2007). Caught bending the A-rule: crystal structures of translesion DNA synthesis with a non-natural nucleotide. *Biochemistry* 46, 10551–10561. doi: 10.1021/bi7008807
- Zahn, K. E., and Doublie, S. (2014). Look Ma, no PCNA: how DNA polymerase epsilon synthesizes long stretches of DNA without a processivity factor. *Nat. Struct. Mol. Biol.* 21, 12–14. doi: 10.1038/nsmb.2749
- Zahn, K. E., Wallace, S. S., and Doublie, S. (2011). DNA polymerases provide a canon of strategies for translesion synthesis past oxidatively generated lesions. *Curr. Opin. Struct. Biol.* 21, 358–369. doi: 10.1016/j.sbi.2011.03.008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 June 2014; paper pending published: 07 July 2014; accepted: 04 August 2014; published online: 25 August 2014.

Citation: Doublie S and Zahn KE (2014) Structural insights into eukaryotic DNA replication. *Front. Microbiol.* 5:444. doi: 10.3389/fmicb.2014.00444

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Doublie and Zahn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field

Sonoko Ishino and Yoshizumi Ishino*

Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka, Japan

Edited by:

Zvi Kelman, University of Maryland,
USA

Reviewed by:

Frank T. Robb, University of Maryland
School of Medicine, USA
Lori Kelman, Montgomery College,
USA

***Correspondence:**

Yoshizumi Ishino, Department of
Bioscience and Biotechnology,
Graduate School of Bioresource and
Bioenvironmental Sciences, Kyushu
University, 6-10-1 Hakozaki,
Higashi-ku, Fukuoka 812-8581, Japan
e-mail: ishino@agr.kyushu-u.ac.jp

DNA polymerase is a ubiquitous enzyme that synthesizes complementary DNA strands according to the template DNA in living cells. Multiple enzymes have been identified from each organism, and the shared functions of these enzymes have been investigated. In addition to their fundamental role in maintaining genome integrity during replication and repair, DNA polymerases are widely used for DNA manipulation *in vitro*, including DNA cloning, sequencing, labeling, mutagenesis, and other purposes. The fundamental ability of DNA polymerases to synthesize a deoxyribonucleotide chain is conserved. However, the more specific properties, including processivity, fidelity (synthesis accuracy), and substrate nucleotide selectivity, differ among the enzymes. The distinctive properties of each DNA polymerase may lead to the potential development of unique reagents, and therefore searching for novel DNA polymerase has been one of the major focuses in this research field. In addition, protein engineering techniques to create mutant or artificial DNA polymerases have been successfully developing powerful DNA polymerases, suitable for specific purposes among the many kinds of DNA manipulations. Thermostable DNA polymerases are especially important for PCR-related techniques in molecular biology. In this review, we summarize the history of the research on developing thermostable DNA polymerases as reagents for genetic manipulation and discuss the future of this research field.

Keywords: thermostability, gene amplification, *in vitro* gene manipulation, Archaea, hyperthermophile

IN THE BEGINNING: TAQ POLYMERASE

DNA polymerase I from *Thermus aquaticus* (Taq polymerase) is the most famous representative enzyme among the thermostable DNA polymerases. Taq polymerase was identified from *T. aquaticus* isolated from Yellowstone National Park in Montana, USA. The report was published by Chien et al. (1976) as her Master's course study. At that time, nobody foresaw how famous this enzyme would later become. In 1985, PCR (polymerase chain reaction) technology using the Klenow fragment of DNA polymerase I from *Escherichia coli* was reported (Saiki et al., 1985). It was easily imagined that a heat-stable DNA polymerase that is not inactivated at the denaturation step from double-stranded to single-stranded DNA would transform this method of gene amplification to a practical technology. Subsequently, a simple and robust PCR method using Taq polymerase was published (Saiki et al., 1988). Due to the heat stability of Taq polymerase, the reaction tube could remain in the incubator after the reaction mixture containing the DNA polymerase was prepared, and only temperature changes were required for PCR. An instrument capable of quick reaction temperature change was developed, and the PCR market opened with a PCR kit (GeneAmp PCR Reagent Kit) and an instrument (Thermal Cycler) provided by Perkin-Elmer Cetus. DNA polymerase from *Thermus thermophilus* (Tth polymerase) was also developed as a commercial product in the early age of the PCR, but a scientific report was only an abstract of ASBMB in 1974 from the Mitsubishi-Kasei Institute of Life Sciences, Japan, where this enzyme was originally identified. A specific property of

Tth polymerase is that it has a distinct reverse transcriptase (RT) activity, and a single-tube RT-PCR method was developed with this enzyme.

At the beginning of the PCR age, Taq polymerase was purified from *T. aquaticus* cells. However, the *pol* gene was soon cloned from the *T. aquaticus* genome and expressed in *E. coli* cells. The native Taq polymerase was replaced by the recombinant Taq polymerase, named AmpliTaq DNA polymerase, in the commercial field. The amount of the recombinant Taq polymerase produced in *E. coli* cells was very low, probably because of the low expression of the *T. aquaticus* gene, which has a high GC content (70%), although the protein quality was improved, as compared to the native Taq polymerase (Lawyer et al., 1989). We successfully constructed an efficient overproduction system by changing the codons around the N-terminal region from the original gene to either the AT-type at the third letter or the optimal codons for *E. coli*. These manipulations improved the production of Taq polymerase more than 10-fold, as compared with the production of AmpliTaq (Ishino et al., 1994). Taq polymerase has been used as the standard enzyme for PCR since its inception. An abundance of PCR data obtained using Taq polymerase has been accumulated, providing a valuable resource for developing new products for useful PCR modifications.

THERMOSTABLE DNA POLYMERASES FROM THERMOPHILES

Thermophilic organisms utilize thermostable DNA polymerases, and therefore, thermophiles became more popular as genetic

resources of DNA polymerases and other enzymes for industrial use. The heat stability of the enzymes is directly related to the temperature, at which the organism thrives. Thermophiles are classified into extreme thermophiles, which grow at temperatures greater than 75°C, and moderate thermophiles, which grow at 55–75°C. The thermostabilities are obviously different between the DNA polymerases from extreme thermophiles and moderate thermophiles as shown in **Figure 1**. Taq polymerase is applicable to PCR; however, the DNA polymerases from the moderately thermophilic *Bacillus* species are not suitable for PCR, because of their insufficient stability. Hyperthermophiles are particular extreme thermophiles that grow optimally at temperatures above 80°C. Most of the hyperthermophilic organisms are Archaea, although some are bacteria, as shown in (**Table 1**). Generally, hyperthermophiles have the potential to provide more heat-stable enzymes than normal thermophiles. Actually, the DNA polymerase from *Pyrococcus furiosus* (Pfu polymerase) is more stable than Taq polymerase (**Figure 1**). Hyperthermophilic archaea became popular not only as sources of useful enzymes for application, but also as interesting model organisms for molecular biology. In the early 1990s, the metabolic phenomena in archaeal cells were just barely understood, and therefore, the molecular biology of Archaea, the third domain of life, became a novel and exciting field.

DNA POLYMERASES FROM HYPERTHERMOPHILES

When choosing thermostable DNA polymerases as reagents for genetic engineering, research scientists generally do not consider the biology of the source organisms. The properties of the obtained enzyme are important, regardless of the source. To obtain a thermostable DNA polymerase, the growth temperature of the thermophile attracts the most attention. *Thermotoga maritima* DNA polymerase was the first commercial product (ULTIMA DNA polymerase) from the hyperthermophilic bacteria. This enzyme has an associated 3'-5' exonuclease activity and thus is expected to perform PCR more accurately with its proofreading

activity. All PCR enzymes from the domain Bacteria are from family A, whose members generally lack 3'-5' exonuclease activity, and ULTIMA DNA polymerase was an exception, like *E. coli* Pol I. In spite of this selling point, ULTIMA DNA polymerase was not a commercial success. One report described no significant differences in the fidelities of the ULTIMA and Taq polymerases, when using optimal buffer conditions for each enzyme, for sequencing purposes (Diaz and Sabino, 1998).

DNA polymerases from the hyperthermophilic archaea were also assessed as PCR enzymes. We cloned the *pol* gene from *P. furiosus* and expressed it in *E. coli* (Uemori et al., 1993). We thought ours would be the first report of the full-length sequence of an archaeal family B DNA polymerase, which had been predicted earlier because of the aphidicolin-sensitive phenotype of a halophile and a methanogen (Forterre et al., 1984; Zabel et al., 1985). However, two papers showing the deduced total amino acid sequences of DNA polymerases from the hyperthermophilic archaea, *Sulfolobus solfataricus* (Pisani et al., 1992) and *Thermococcus litoralis* (Perler et al., 1992) were published during the preparation of our manuscript (Uemori et al., 1993). All these reports clearly showed that the archaeal DNA polymerases have sequences similar to the

Table 1 | Representative hyperthermophiles.

Species	Growth conditions		
	Temperature		
	Min. temp to Max. temp (°C)	Opt. temp (°C)	Aerobic(ae)/ Anaerobic(an)
Archaea			
Crenarchaeota			
<i>Acidianus infernus</i>	60–95	88	ae/an
<i>Sulfolobus acidocaldarius</i>	60–85	75	ae
<i>Pyrobaculum islandicum</i>	74–103	100	an
<i>Thermoproteus tenax</i>	70–97	88	an
<i>Desulfurococcus mobilis</i>	70–95	85	an
<i>Aeropyrum pernix</i>	70–100	90	ae
<i>Ignicoccus islandicus</i>	65–100	90	an
<i>Pyrolobus fumarii</i>	90–113	106	ae/an
Euryarchaeota			
<i>Pyrococcus furiosus</i>	70–105	100	an
<i>Thermococcus kodakarensis</i>	60–100	85	an
<i>Archaeoglobus fulgidus</i>	60–95	83	an
<i>Methanopyrus kandleri</i>	90–122	105	an
<i>Methanothermus sociabilis</i>	65–97	88	an
<i>Methanococcus igneus</i>	45–91	88	an
Bacteria			
<i>Thermotoga maritima</i>	55–90	80	an
<i>Aquifex pyrophilus</i>	67–95	85	ae

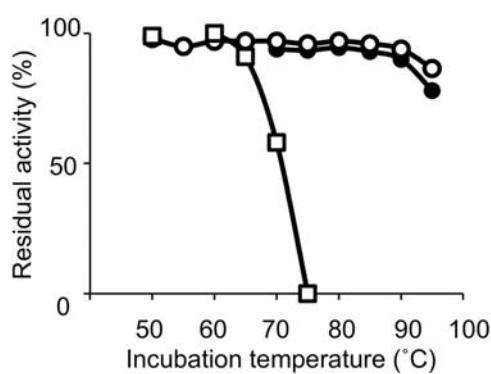


FIGURE 1 | Heat resistance of the DNA polymerases. Residual DNA polymerase activities after incubation at the indicated temperature for 30 min were plotted. DNA polymerases from *Pyrococcus furiosus* (open circles), *Thermus aquaticus* (closed circles), and *Bacillus caldotenax* (open squares) were used as representatives from hyperthermophiles, extreme thermophiles, and moderate extremophiles, respectively.

eukaryotic replicative DNA polymerases, Pol α , δ , and ϵ (family B). It is also interesting that the *T. litoralis pol* has inteins that must be spliced out after translation (Perler et al., 1992). Thereafter, many cases of DNA polymerases containing various pattern of inteins, inserted in motifs A, B, and C, were discovered (Perler, 2002). The fidelity of DNA synthesis *in vitro* is markedly affected by the reaction condition. However, the archaeal family B enzymes generally perform more accurate DNA synthesis as compared with Taq polymerase (Cariello et al., 1991; Ling et al., 1991; Lundberg et al., 1991; Mattila et al., 1991), suggesting that the strong 3'-5' exonuclease activities of the hyperthermophilic family B polymerase *in vitro* affect the fidelity of PCR.

DEVELOPMENT OF LA-PCR

DNA polymerases are classified into seven families based on the amino acid sequence similarity (Figure 2). To date, the enzymes utilized for genetic engineering have been only from families A and B among them. Taq polymerase from family A has strong extension ability and performs efficient amplification of the target DNA. However, their fidelity is low. On the other hand, the Pfu polymerase from family B performs highly accurate PCR amplification, but their extension rate is slow and a long extension time is required for each cycle of PCR. Therefore, a method was required for the accurate PCR amplification of long DNA regions. One simple idea that researchers considered trying was to combine one enzyme each from family A and family B in a single PCR reaction mixture. However, the actual PCR performance was not so simple, and persevering trials were necessary to find suitable conditions to develop a long and accurate (LA) PCR system. The amplification of a ~35 kb DNA fragment from λ phage genomic DNA was successfully accomplished in 1994, by the mixture of KlenTaq1 (N-terminal deletion mutant of Taq polymerase) and an archaeal family B DNA polymerase with 3'-5' exonuclease activity (Barns, 1994). Subsequently, commercial products for LA-PCR were rapidly developed by several manufactures and LA-PCR technology became popular throughout the world.

FAST AND HIGHLY ACCURATE PCR BY AN ARCHAEOAL FAMILY B DNA POLYMERASE

A family B DNA polymerase from the hyperthermophilic archaeon, *Thermococcus kodakarensis* (this strain was originally named *Pyrococcus kodakaraensis* KOD1), was identified and

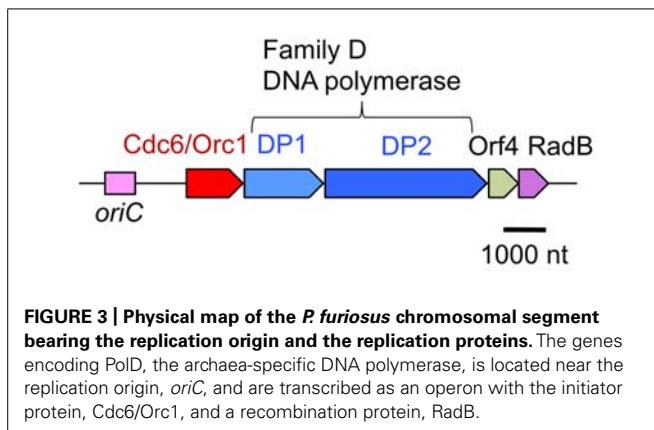
applied to PCR (Takagi et al., 1997). This enzyme has the typical amino acid sequence of the archaeal family B enzymes, but it showed a high extension rate while maintaining high fidelity, and therefore, the commercial product, KOD DNA polymerase (KOD Pol), was developed and became popular as a PCR enzyme. Commercial products related to KOD Pol, including a hot start kit with a monoclonal antibody and an LA-PCR kit with a mixture of the wild type and 3'-5' exonuclease-deficient mutant of this enzyme, were subsequently developed by the manufacturers. The underlying reason why this family B enzyme shows high extension speed is interesting. Comparisons of the crystallographic structures and amino acid sequences of KOD Pol with other archaeal family B enzymes revealed the logical explanation for the efficient extension ability of this enzyme. Many basic residues are located around the active site in the finger domain of KOD Pol. In addition, many Arg residues are located at the forked point, which is the predicted as the junction of the template binding region and the editing cleft. This unique structure may stabilize the melted DNA structure at the forked point, resulting in high PCR performance (Hashimoto et al., 2001).

BASIC RESEARCH ON ARCHAEOAL DNA POLYMERASES

Research on DNA polymerases in hyperthermophilic archaea is motivated by not only industrial applications, but also basic molecular biology, to elucidate the molecular mechanisms of genetic information processing systems at extremely hot temperatures. To identify all of the DNA polymerases in the archaeal cell, we tried to separate the DNA polymerase activities in the total cell extract of *P. furiosus*. Three major fractions showed nucleotide incorporation activity after anion exchange column chromatography (Resource Q column, GE Healthcare; Imamura et al., 1995). In addition to the further purification of each fraction, the screening of the DNA polymerase activity from the heat-stable protein library, made from *E. coli* cell extracts containing *P. furiosus* DNA fragments, revealed a new DNA polymerase gene (Uemori et al., 1997). The new DNA polymerase consisted of two proteins, the small and large subunits, and we named it DP1 and DP2. These two proteins are strictly required for both 5'-3' polymerizing and 3'-5' exonucleolytic activities *in vitro*. The genes encoding DP1 and DP2 are located in tandem on the *P. furiosus* genome and form an operon. Interestingly, this operon has a total of five genes, including a gene encoding a eukaryotic Cdc6/Orc1 protein (important for initiation of DNA replication) and a gene encoding a Rad51-like protein (involved in homologous recombination in Eukarya), in addition to DP1 and DP2 (Figure 3; Uemori et al., 1997). This was the first report of a eukaryotic-like initiator protein for DNA replication in Archaea. The amino acid sequences of DP1 and DP2 are not similar to those of any other DNA polymerases. After the discovery of this DNA polymerase, the total genome sequence of *Methanococcus jannaschii* was published as the first complete archaeal genome (Bult et al., 1996). One of the topics of this report was that only one DNA polymerase (family B) was found in the deduced amino acid sequences, in contrast to the three DNA polymerases, PolI, II, and III, in *E. coli* and several DNA polymerases in eukaryotic cells (Gray, 1996). We searched for homologous sequences of DP1 and DP2 in the *M. jannaschii* genome, and found them. The two genes

	Family						
	A	B	C	D	E	X	Y
Bacteria	I ■	II ■	III ■				IV ■ V ■
Archaea		B1 ■ B3 ■		D ■	E ■		Y ■
Eukarya	γ θ ■ ■	α δ ϵ ζ ■ ■ ■ ■			β λ μ σ ■ ■ ■ ■		η ι κ ■ ■ ■

FIGURE 2 | Distribution of DNA polymerases in the three domains of life. The names of DNA polymerases vary, depending on the domains. Only DNA polymerases with *in vitro* activity, if applicable, are shown. Eukaryotic Poly is from mitochondria and archaeal PolE is a plasmid-encoded enzyme.



were not present in tandem, but were located separately on the genome. We cloned and expressed them in *E. coli*, and demonstrated their polymerase and exonuclease activities *in vitro*. With this report, DP1 and DP2 became recognized as a novel archaeal DNA polymerase (Ishino et al., 1998). Three more total genome sequences were subsequently reported, and the genes for DP1 and DP2 were found in all them. Thus, this new DNA polymerase became more generally found in Archaea (Cann et al., 1998). Due to the lack of sequence homology to other DNA polymerases, we proposed a new family, family D, for this enzyme (Cann and Ishino, 1999).

In parallel to the identification of DNA polymerase activities in the cell extract of *P. furiosus*, we amplified a gene fragment for the family B DNA polymerase from the genomic DNA of *Pyrodictium occultum*, which grows at 105°C, in an attempt to find a more heat-stable DNA polymerase than that from *P. furiosus*. By using a set of mixed primers based on the conserved sequences of motifs A and C in the family B DNA polymerase, a single band was amplified. However, two different fragments were found after the cloning and sequencing of the PCR product. The full-length sequences of both *pol*-like genes were cloned from the *P. occultum* genome by the primer walking method, and they were expressed in *E. coli*. Both of the gene products exhibited the heat stable DNA polymerase activity (Uemori et al., 1995). Unfortunately, the performance of these two enzymes in PCR was not better than Pfu polymerase, and we discontinued further research on them. However, this was the first report that an archaeal cell has two different family B DNA polymerases. It was an exciting discovery because three family B DNA polymerases, Pol α , Pol δ , and Pol ϵ , were known in eukarya, and so we proposed that plural family B enzymes were a common feature between Archaea and Eukarya. However, there is only one gene encoding a family B DNA polymerase in the *M. jannaschii* genome as described above. We subsequently found two family B DNA polymerases in *Aeropyrum pernix* (Cann et al., 1999), and thus the presence of two family B enzymes is not special for *Pyrodictium*, but is more general in Archaea. In the early stages of the total genome sequences, all sequences were from Euryarchaeota (*Archaeoglobus fulgidus*, *Methanothermobacter thermautotrophicus*, *Pyrococcus horikoshii*) and the determination of the genome sequence of a crenarchaeal organism was delayed until that of *A. pernix* was reported (Kawarabayasi et al., 1999). Taken together

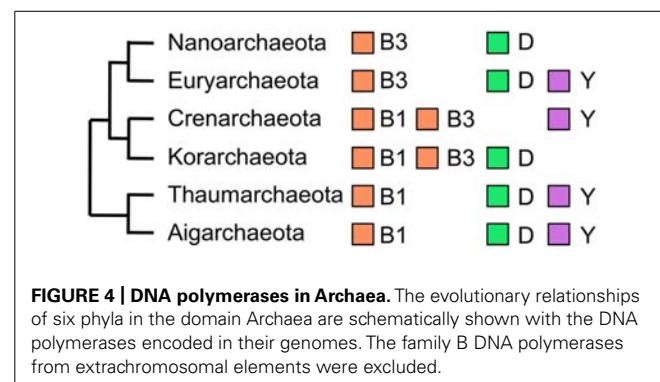
with the new knowledge at that time, it was predicted that euryarchaeal organisms have one DNA polymerase each from family B and family D, respectively, and crenarchaeal organisms have at least two family B enzymes in the cell. This overview of the distribution of DNA polymerases in Archaea is generally correct as shown in (Figure 4), which displays DNA polymerases in the archaeal phyla (subdomains) including newly proposed phyla from recent ecological research.

All of the original biochemical data for *P. furiosus* PolD from our group, including thermostability, strong primer extension and 3'-5' exonuclease activity, showed that PolD is a suitable enzyme for PCR (Ishino and Ishino, 2001). However, PolD has not been commercially developed. Recent analysis of *Pyrococcus abyssi* PolD revealed that it is a suitable PCR enzyme (Killelea et al., 2014). On the contrary, PolD from *Thermococcus* sp 9°N does not have any advantages as compared with the current commercially available PCR enzymes (Greenough et al., 2014).

PROTEIN ENGINEERING OF THERMOSTABLE DNA POLYMERASES

Once PCR technology was established, efforts to improve PCR performance were pursued. At the early stage, hot start PCR was one of the big improvements for the specific amplification. An antibody against Taq polymerase was used to suppress its enzyme activity by specific antigen–antibody binding at the low temperature, and when PCR started from the denaturing temperature at more than 90°C, the antibody became separated from the enzyme by heat denaturation. This hot start PCR method is generally effective to prevent non-specific amplification. For this purpose, another idea was tested. A chemical modification of Taq polymerase inactivated its enzymatic activity at low temperatures, but the modification can be released by high temperature resulting in activation of Taq polymerase to start PCR. This temperature-dependent reversible modification of the Taq protein led to the commercial product, AmpliTaq Gold, as the hot start PCR enzyme. Alternatively, a cold-sensitive Taq polymerase with markedly reduced activity at 37°C, as compared with the wild type enzyme, was produced by site-directed mutagenesis, and this mutant is suitable for hot start PCR (Kermekchiev et al., 2003).

Taq polymerase is a family A enzyme, and is applicable to practical dideoxy sequencing. However, the output of the sequencing data was not ideal as compared with that from T7 DNA polymerase



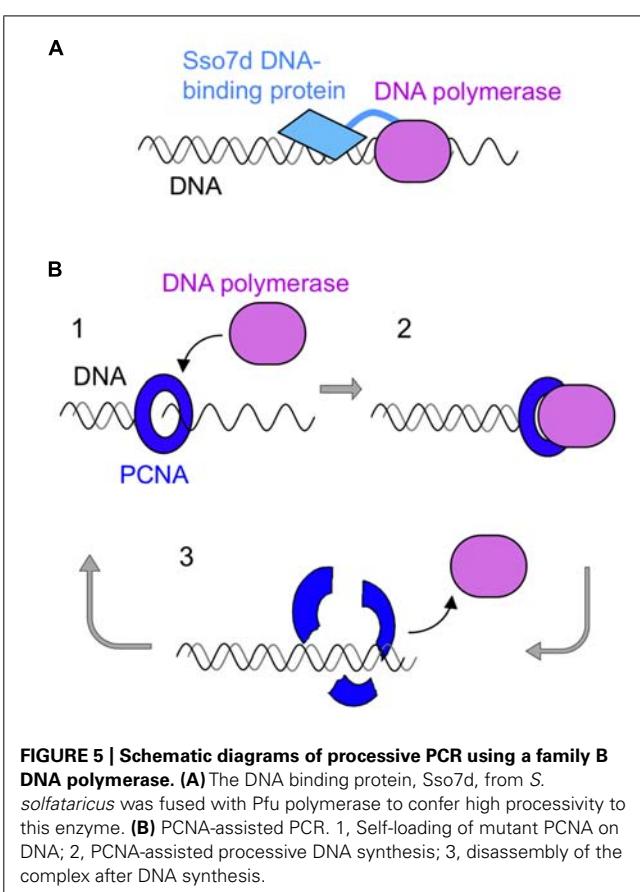
(known commercially as Sequenase; see below). An ingenious protein engineering strategy produced a mutant Taq polymerase that is more suitable for dideoxy sequencing than the wild type Taq polymerase. *E. coli* PolII and Taq polymerase discriminate deoxy- and dideoxynucleotide as substrates for the incorporation into the DNA strand, and therefore, an excess amount (50 to 1000-fold) of dideoxynucleotides must be present in the reaction mixture to stop DNA strand synthesis by their incorporation. For this property, the strength of each signal is not uniform, but is distinctly unbalanced. However, T7 DNA polymerase equally incorporates deoxynucleotides and dideoxynucleotides, and therefore, it is easy to adjust the reaction conditions to provide very clear signals (Tabor and Richardson, 1990). A mutant T7 DNA polymerase lacking the 3'-5' exonuclease activity was developed as a commercial product, named Sequenase. A detailed comparison of *E. coli* Pol I and T7 polymerase revealed one amino acid that discriminates deoxy- and dideoxynucleotides, resulting in the successful conversions of the properties from PolII to T7 and T7 to PolII (Tabor and Richardson, 1995). This work was applied to Taq polymerase and a modified Taq with F667Y, which endows Taq with T7-type substrate recognition, was created (Tabor and Richardson, 1995). This enzyme was called Thermosequenase, and it became popular as the standard enzyme for the fluorescently labeled sequencing method (Reeve and Fuller, 1995).

Another target for the creation of a new enzyme by mutagenesis is an enzyme that is more resistant to PCR inhibitors in blood or soil, such as hemoglobin and humic acid. A mutant Taq DNA polymerase with enhanced resistance to various inhibitors, including whole blood, plasma, hemoglobin, lactoferrin, serum IgG, soil extracts, and humic acid, was successfully created by site-directed mutagenesis (Kermekchiev et al., 2009). The molecular breeding of *Thermus* DNA polymerases by using a direct evolution technique, compartmentalized self-replication (CSR; Ghadessy et al., 2001), also generated a PCR enzyme with striking resistance to a broad spectrum of inhibitors with highly divergent compositions, including humic acid, bone dust, coprolite, peat extract, clay-rich soil, cave sediment, and tar (Baar et al., 2011). Furthermore, enzymes with a broad substrate specificity spectrum, which are thus useful for the amplification of ancient DNA containing numerous lesions, were also obtained by the CSR technique (Ghadessy et al., 2004; d'Abbadie et al., 2007). Mutational studies of the O-helix of Taq DNA polymerase produced enzymes with reduced fidelity, which may be useful for error-prone PCR (Suzuki et al., 1997, 2000; Tosaka et al., 2001).

One successful strategy to produce improved DNA polymerases is the “domain tagging.” For example, new DNA polymerases were created by the flexible attachment of helix–hairpin–helix (HhH) domains of *Methanopyrus kandleri* topoisomerase V to the catalytic domains of Taq polymerase and Pfu polymerases. HhH is a widespread motif and generally functions on sequence-nonspecific DNA binding. These hybrid enzymes increased thermostability and became more resistant to salt and several inhibitors such as phenol, blood, and DNA intercalating dyes (Pavlov et al., 2002). This tagging strategy was also applied to φ29 DNA polymerase (de Vega et al., 2010) and *Bacillus stearothermophilus* DNA polymerase (Pavlov et al., 2012). Another

successful example of the tagging strategy is the creation of commercial product “Phusion DNA polymerase” (Figure 5). This is a fusion protein of Pfu DNA polymerase and a DNA binding protein, Sso7d, from *S. solfataricus* (Wang et al., 2004). Sso7d has strong affinity to DNA, and it retains the fused Pfu DNA polymerase on the DNA once it starts DNA synthesis along with the template DNA strand. Phusion DNA polymerase compensates for the low extension rate of Pfu DNA polymerase while maintaining its high fidelity. This enzyme shows very high processivity and accurate PCR performance, and is now widely used.

Another idea to improve the processivity of the archaeal family B DNA polymerases was to use PCNA (proliferating cell nuclear antigen) as a processivity factor. The ring-shaped PCNA encircles the DNA strand and slides on it, and various binding proteins are attached to PCNA (Pan et al., 2011). DNA polymerase is a typical PCNA binding protein and it is connected to the DNA strand by PCNA during strand synthesis. This is why DNA polymerase shows highly processive DNA synthesis in the presence of PCNA. Based on this property of PCNA, scientists have searched for a thermostable PCNA for PCR with DNA polymerase. However, PCNA has not yet been successfully used for PCR. Unexpectedly, PCR is inhibited, rather than stimulated, in the presence of PCNA. We developed a PCNA-assisted PCR method, which is highly processive PCR with high fidelity, by using a mutant PCNA. Originally, we determined the crystal structure of *P. furiosus* PCNA (PfuPCNA; Matsumiya et al., 2001), and our continued research



revealed that the intermolecular ion pairs between the protomers of PfuPCNA contributed to its ring stability (Matsumiya et al., 2003), which was greatly affected by the ionic strength of the solution. Mutations of the amino acid residues involved in the ion pairs clearly decreased its ring stability, but unexpectedly, a less stable mutant PfuPCNA enhanced the primer extension reaction of Pfu DNA polymerase *in vitro* (Matsumiya et al., 2003). Therefore, we applied the mutant PfuPCNA to PCR and successfully amplified DNA fragments up to 15 kbp with a markedly shorter reaction time, by Pfu DNA polymerase in the presence of a PfuPCNA mutant under conditions where Pfu DNA polymerase alone did not function (Ishino et al., 2012; Kawamura et al., 2012). This PCNA-assisted PCR (**Figure 5**) is also a successful example of processive PCR with high accuracy.

Because of the high sensitivity of PCR, very small amounts of carry-over contaminants from previous PCRs are considered to be one of the major sources of false positive results. The most common strategy to prevent carry-over contamination is to replace dTTP with dUTP during PCR amplification, thereby producing DNA containing uracil. Prior to initiating PCR, the PCR mixture is treated with Uracil-DNA glycosylase (UNG). During the initial denaturation step temperature is elevated to 95°C, resulting in cleavage of apyrimidinic sites and fragmentation of carry-over DNA. One problem of the archaeal family B DNA polymerase to be used for this carry-over prevention is that they specifically interact with uracil and hypoxanthine, which stalls their progression on DNA template strands (Connolly, 2009). The crystal structure of the DNA polymerase revealed that read-ahead recognition occurs by an interaction with the deaminated bases in an N-terminal binding pocket that is specifically found in the archaeal family B DNA polymerases (Fogg et al., 2002). Due to this specific recognition of uracil, the archaeal family B DNA polymerases, including Pfu DNA polymerase and KOD DNA polymerase, are not suitable for carry-over prevention PCR. To conquer this defect, a point mutation (V98Q) was introduced into Pfu polymerase. This mutant enzyme is completely unable to recognize uracil, while its DNA polymerase activity is unaffected (Fogg et al., 2002; Firbank et al., 2008). Therefore, this mutant Pfu polymerase is useful for the carry-over prevention PCR. It is also useful for amplification of uracil-containing DNA, such as damaged DNA and bisulfite-converted DNA for epigenetic analysis.

FUTURE PERSPECTIVES

Polymerase chain reaction initiated a revolution in molecular biology, and is now used daily not only in research, but also in the general human society. PCR is a complete technology, but more powerful and reliable enzymes for PCR are still desired. Notably, an enzyme with faster, longer, and more efficient extension ability, as compared to the properties of the current commercial products, will contribute to further improvements in PCR technology. In addition to these basic abilities, DNA polymerases that can incorporate various modified nucleotides, which are useful for highly sensitive labeling, are valuable for single molecule analysis. Mutations of the DNA polymerase itself, by site-specific or random mutagenesis, are effective ways to create modified enzymes with improved PCR performance or specific properties

for *in vitro* DNA manipulations. An artificial evolution procedure also has attracted a great deal of attention, for the creation of DNA polymerases with novel activities (Brakmann, 2005; Henry and Romesberg, 2005; Holmberg et al., 2005; Ong et al., 2006). Our strategy of using environmental DNA as a genetic resource also works well to investigate the structure–function relationships of DNA polymerases. The region corresponding to the active center of the DNA polymerizing reaction, in the structural genes of Taq polymerase and Pfu polymerase, was substituted with PCR fragments amplified from DNAs within soil samples from various locations in Japan. The chimeric *pol* genes were constructed within the expression plasmids for the Taq and Pfu polymerases in *E. coli*. The chimeric enzymes thus produced, exhibited DNA polymerase activities with different properties (Matsukawa et al., 2009). The main focus for the future development of DNA polymerases is not on versatile enzymes, but rather on specialized enzymes suitable for individual purposes, including whole genome amplification, rapid detection of short DNA, new sequencing technologies, etc. Continued research on DNA polymerases may facilitate the invention of new genetic analysis technologies that are completely different from PCR or PCR-related techniques. The isothermal amplification without temperature cycling is more convenient and practical than PCR, and development of this type of technique has been actively performed (Gill and Ghaemi, 2008). Several methods practically utilized now are based on the strand displacement (SD) activity of the DNA polymerases. DNA polymerases from φ29 bacteriophage and *B. stearothermophilus* are the representative enzymes for the SD activity. A whole genome amplification using the SD activity of φ29 DNA polymerase is now especially useful for single cell analysis. Alternatively, helicase was applied for the dissociation of the double-stranded DNA from an idea to mimic DNA replication *in vivo* (Vincent et al., 2004). Although the helicase-dependent amplification (HDA) technique has not been practically used (Jeong et al., 2009), brushing up this technique may generate a powerful tool for genetic engineering.

ACKNOWLEDGMENTS

The writing of this review article was supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (grant number 26242075 to Yoshizumi Ishino).

REFERENCES

- Baar, C., d'Abbadie, M., Vaisman, A., Arana, M. E., Hofreiter, M., Woodgate, R., et al. (2011). Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.* 39:e51. doi: 10.1093/nar/gkq1360
- Barns, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2216–2220. doi: 10.1073/pnas.91.6.2216
- Brakmann, S. (2005). Directed evolution as a tool for understanding and optimizing nucleic acid polymerase function. *Cell. Mol. Life Sci.* 62, 2634–2646. doi: 10.1007/s0018-005-5165-5
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., et al. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073. doi: 10.1126/science.273.5278.1058
- Cann, I. K. O., Ishino, S., Nomura, N., Sako, Y., and Ishino, Y. (1999). Two family B DNA polymerases in *Aeropyrum pernix*, an obligate aerobic hyperthermophilic crenarchaeote. *J. Bacteriol.* 181, 5984–5992.
- Cann, I. K. O., and Ishino, Y. (1999). Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* 152, 1249–1267.

- Cann, I., Komori, K., Toh, H., Kanai, S., and Ishino, Y. (1998). A heterodimeric DNA polymerase: evidence that members of the euryarchaeota possess a novel DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14250–14255. doi: 10.1073/pnas.95.24.14250
- Cariello, N. F., Swenberg, J. A., and Skopek, T. R. (1991). Fidelity of *Thermococcus litoralis* DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* 19, 4193–4198. doi: 10.1093/nar/19.15.4193
- Chien, A., Edgar, D. B., and Trella, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* 127, 1550–1557.
- Connolly, B. A. (2009). Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochem. Soc. Trans.* 37, 65–68. doi: 10.1042/BST0370065
- d'Abbadie, M., Hofreiter, M., Caisman, A., Loakes, D., Gasparutto, D., Cadet, J., et al. (2007). Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* 25, 939–943. doi: 10.1038/nbt1321
- de Vega, M., Lázaro, J. M., Mencía, M., Blanco, L., and Salas, M. (2010). Improvement of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16506–16511. doi: 10.1073/pnas.1011428107
- Diaz, R. S., and Sabino, E. C. (1998). Accuracy of replication in the polymerase chain reaction. Comparison between *Thermotoga maritima* DNA polymerase and *Thermus aquaticus* DNA polymerase. *Braz. J. Med. Biol. Res.* 31, 1239–1242. doi: 10.1590/S0100-879X1998001000001
- Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J., and Connolly, B. A. (2008). Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.* 381, 529–539. doi: 10.1016/j.jmb.2008.06.004
- Fogg, M. G., Pearl, L. H., and Connolly, B. A. (2002). Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* 9, 922–927. doi: 10.1038/nsb867
- Forterre, P., Elie, C., and Kohiyama, M. (1984). Aphidicolin inhibits growth and DNA synthesis in halophilic archaeabacteria. *J. Bacteriol.* 159, 800–802.
- Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4552–4557. doi: 10.1073/pnas.071052198
- Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., et al. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* 22, 755–779. doi: 10.1038/nbt974
- Gill, P., and Ghaemi, A. (2008). Nucleic acid isothermal amplification technologies. *Nucleosides Nucleotides Nucleic Acids* 27, 224–243. doi: 10.1080/15257770701845204
- Gray, M. (1996). The third form of life. *Nature* 383:299. doi: 10.1038/383299a0
- Greenough, L., Menin, J. F., Desai, N. S., Kelman, Z., and Gardner, A. F. (2014). Characterization of family D DNA polymerase from *Thermococcus* sp. 9°N. *Extremophiles* 18, 653–664. doi: 10.1007/s00792-014-0646-9
- Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., et al. (2001). Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J. Mol. Biol.* 306, 469–477. doi: 10.1006/jmbi.2000.4403
- Henry, A. A., and Romesberg, F. E. (2005). The evolution of DNA polymerases with novel activities. *Curr. Opin. Biotechnol.* 16, 370–377. doi: 10.1016/j.copbio.2005.06.008
- Holmberg, R. C., Henry, A. A., and Romesberg, F. E. (2005). Directed evolution of novel polymerases. *Biomol. Eng.* 22, 39–49. doi: 10.1016/j.biomed.2004.12.001
- Imamura, M., Uemori, T., Kato, I., and Ishino, Y. (1995). A non-alpha-like DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biol. Pharm. Bull.* 18, 1647–1652. doi: 10.1248/bpb.18.1647
- Ishino, S., Kawamura, A., and Ishino, Y. (2012). Application of PCNA to processive PCR by reducing the stability of its ring structure. *J. Jpn. Soc. Extremophiles* 11, 19–25.
- Ishino, Y., and Ishino, S. (2001). Novel DNA polymerases from Euryarchaeota. *Methods Enzymol.* 334, 249–260. doi: 10.1016/S0076-6879(01)34473-7
- Ishino, Y., Komori, K., Cann, I. K., and Koga, Y. (1998). A novel DNA polymerase family found in Archaea. *J. Bacteriol.* 180, 2232–2236.
- Ishino, Y., Ueno, T., Miyagi, M., Uemori, T., Imamura, M., Tsunashima, S., et al. (1994). Overproduction of *Thermus aquaticus* DNA polymerase and its structural analysis by ion-spray mass spectrometry. *J. Biochem.* 116, 1019–1024.
- Jeong, Y. J., Park, K., and Kim, D. E. (2009). Isothermal DNA amplification in vitro: the helicase-dependent amplification system. *Cell. Mol. Life Sci.* 66, 3325–3336. doi: 10.1007/s00018-009-0094-3
- Kawamura, A., Ishino, Y., and Ishino, S. (2012). Biophysical analysis of PCNA from *Pyrococcus furiosus*. *J. Jpn. Soc. Extremophiles* 11, 12–18.
- Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., et al. (1999). Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* 6, 83–101, 145–152. doi: 10.1093/dnares/6.2.145
- Kermekchiev, M. B., Kirillova, L. I., Vail, E. E., and Barnes, W. M. (2009). Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Res.* 37:e40. doi: 10.1093/nar/gkn1055
- Kermekchiev, M. B., Tzakov, A., and Barnes, W. M. (2003). Cold-sensitive mutants of Taq DNA polymerase provide a hot start for PCR. *Nucleic Acids Res.* 31, 6139–6147. doi: 10.1093/nar/gkg813
- Killelea, T., Ralec, C., Bossé, A., and Henneke, G. (2014). PCR performance of a thermostable heterodimeric archaeal DNA polymerase. *Front. Microbiol.* 5:195. doi: 10.3389/fmicb.2014.00195
- Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989). Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *J. Biol. Chem.* 264, 6427–6437.
- Ling, L. L., Keohavong, P., Dias, C., and Thilly, W. G. (1991). Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and Vent DNA polymerases. *PCR Methods Appl.* 1, 63–69. doi: 10.1101/gr.1.1.63
- Lundberg, K. S., Shoemaker, D. D., Adams, M. W., Short, J. M., Sorge, J. A., and Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* 108, 1–6. doi: 10.1016/0378-1119(91)90480-Y
- Matsukawa, H., Yamagami, T., Kawarabayasi, Y., Miyashita, Y., Takahashi, M., and Ishino, Y. (2009). A useful strategy to construct DNA polymerases with different properties by using genetic resources from environmental DNA. *Genes Genet. Syst.* 84, 3–13. doi: 10.1266/ggs.84.3
- Matsumiya, S., Ishino, S., Ishino, Y., and Morikawa, K. (2003). Intermolecular ion pairs maintain toroidal structure of *Pyrococcus furiosus* PCNA. *Protein Sci.* 12, 823–831. doi: 10.1110/ps.0234503
- Matsumiya, S., Ishino, Y., and Morikawa, K. (2001). Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen of *Pyrococcus furiosus*. *Protein Sci.* 10, 17–23. doi: 10.1110/ps.36401
- Mattila, P., Korpela, J., Tenkanen, T., and Pitkänen, K. (1991). Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase – an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.* 19, 4967–4973. doi: 10.1093/nar/19.18.4967
- Ong, J. L., Loakes, D., Jaroslawski, S., Too, K., and Holliger, P. (2006). Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* 361, 537–550. doi: 10.1016/j.jmb.2006.06.050
- Pan, M., Kelman, L. M., and Kelman, Z. (2011). The archaeal PCNA proteins. *Biochem. Soc. Trans.* 39, 20–24. doi: 10.1042/BST0390020
- Pavlov, A. R., Belova, G. I., Kozyavkin, S. A., and Slesarev, A. I. (2002). Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13510–13515. doi: 10.1073/pnas.202127199
- Pavlov, A. R., Pavlova, N. V., Kozyavkin, S. A., and Slesarev, A. I. (2012). Cooperation between catalytic and DNA binding domains enhances thermostability and supports DNA synthesis at higher temperatures by thermostable DNA polymerases. *Biochemistry* 51, 2032–2043. doi: 10.1021/bi2014807
- Perler, F. B. (2002). InBase, the InTein Database. *Nucleic Acids Res.* 30, 383–384. doi: 10.1093/nar/30.1.383
- Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., et al. (1992). Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5577–5581. doi: 10.1073/pnas.89.12.5577
- Pisani, F. M., De Martino, C., and Rossi, M. (1992). A DNA polymerase from the archaeon *Sulfolobus solfataricus* shows sequence similarity to family B DNA polymerases. *Nucleic Acids Res.* 20, 2711–2716. doi: 10.1093/nar/20.11.2711
- Reeve, M. A., and Fuller, C. W. (1995). A novel thermostable polymerase for DNA sequencing. *Nature* 376, 796–797. doi: 10.1038/376796a0
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., et al. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491. doi: 10.1126/science.2448875

- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354. doi: 10.1126/science.2999980
- Suzuki, M., Avicola, A. K., Hood, L., and Loeb, L. A. (1997). Low fidelity mutants in the O-helix of *Thermus aquaticus* DNA polymerase I. *J. Biol. Chem.* 272, 11228–11235. doi: 10.1074/jbc.272.17.11228
- Suzuki, M., Yoshida, S., Adman, E. T., Blank, A., and Loeb, L. A. (2000). *Thermus aquaticus* DNA polymerase I mutants with altered fidelity. Interacting mutations in the O-helix. *J. Biol. Chem.* 275, 32728–32735. doi: 10.1074/jbc.M000097200
- Tabor, S., and Richardson, C. C. (1990). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Effect of pyrophosphorolysis and metal ions. *J. Biol. Chem.* 265, 8322–8328.
- Tabor, S., and Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6339–6343. doi: 10.1073/pnas.92.14.6339
- Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., et al. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* 63, 4504–4510.
- Tosaka, A., Ogawa, M., Yoshida, S., and Suzuki, M. (2001). O-helix mutant T664P of *Thermus aquaticus* DNA polymerase I: altered catalytic properties for incorporation of incorrect nucleotides but not correct nucleotides. *J. Biol. Chem.* 276, 27562–27567. doi: 10.1074/jbc.M010635200
- Uemori, T., Ishino, Y., Doi, H., and Kato, I. (1995). The hyperthermophilic archaeon *Pyrodictium occultum* has two α -like DNA polymerases. *J. Bacteriol.* 177, 2164–2177.
- Uemori, T., Ishino, Y., Toh, H., Asada, K., and Kato, I. (1993). Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.* 21, 259–265. doi: 10.1093/nar/21.2.259
- Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, Y. (1997). A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells* 2, 499–512. doi: 10.1046/j.1365-2443.1997.1380336.x
- Vincent, M., Xu, Y., and Kong, H. (2004). Helicase-dependent isothermal DNA amplification. *EMBO Rep.* 5, 795–800. doi: 10.1038/sj.embo.7400200
- Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* 32, 1197–1207. doi: 10.1093/nar/gkh271
- Zabel, H. P., Fischer, H., Holler, E., and Winter, J. (1985). In vivo and in vitro evidence for eucaryotic α -type DNA-polymerases in methanogens. Purification of the DNA-polymerase of *Methanococcus vannielii*. *Syst. Appl. Microbiol.* 6, 111–118. doi: 10.1016/S0723-2020(85)80042-4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2014; accepted: 15 August 2014; published online: 29 August 2014.

Citation: Ishino S and Ishino Y (2014) DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field. *Front. Microbiol.* 5:465. doi: 10.3389/fmicb.2014.00465

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Ishino and Ishino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



DNA polymerase hybrids derived from the family-B enzymes of *Pyrococcus furiosus* and *Thermococcus kodakarensis*: improving performance in the polymerase chain reaction

Ashraf M. Elshawadfy[†], Brian J. Keith, H'Ng Ee Ooi, Thomas Kinsman, Pauline Heslop and Bernard A. Connolly*

Institute of Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne, UK

Edited by:

Zvi Kelman, University of Maryland, USA

Reviewed by:

Juergen Reichardt, James Cook University, Australia

David L. Bernick, University of California, Santa Cruz, USA

***Correspondence:**

Bernard A. Connolly, Institute of Cell and Molecular Biosciences (ICAMB), University of Newcastle, The Medical School, Framlington Place, Newcastle upon Tyne, NE24HH, UK
e-mail: bernard.connolly@ncl.ac.uk

†Present address:

Ashraf M. Elshawadfy, Botany Department, Zagazig University, Zagazig, Egypt

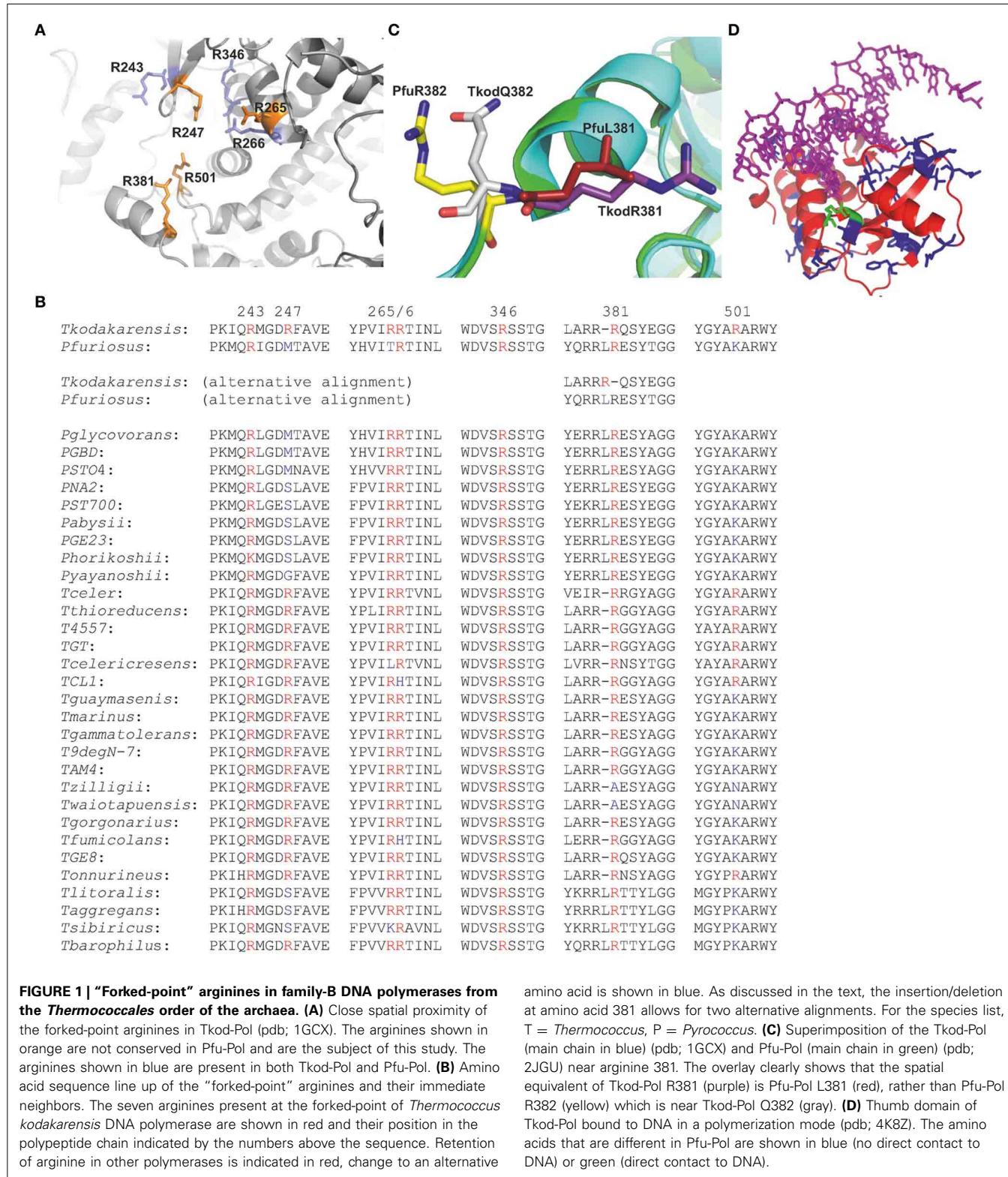
The polymerase chain reaction (PCR) is widely applied across the biosciences, with archaeal Family-B DNA polymerases being preferred, due to their high thermostability and fidelity. The enzyme from *Pyrococcus furiosus* (Pfu-Pol) is more frequently used than the similar protein from *Thermococcus kodakarensis* (Tkod-Pol), despite the latter having better PCR performance. Here the two polymerases have been comprehensively compared, confirming that Tkod-Pol: (1) extends primer-templates more rapidly; (2) has higher processivity; (3) demonstrates superior performance in normal and real time PCR. However, Tkod-Pol is less thermostable than Pfu-Pol and both enzymes have equal fidelities. To understand the favorable properties of Tkod-Pol, hybrid proteins have been prepared. Single, double and triple mutations were used to site arginines, present at the "forked-point" (the junction of the exonuclease and polymerase channels) of Tkod-Pol, at the corresponding locations in Pfu-Pol, slightly improving PCR performance. The Pfu-Pol thumb domain, responsible for double-stranded DNA binding, has been entirely replaced with that from Tkod-Pol, again giving better PCR properties. Combining the "forked-point" and thumb swap mutations resulted in a marked increase in PCR capability, maintenance of high fidelity and retention of the superior thermostability associated with Pfu-Pol. However, even the arginine/thumb swap mutant falls short of Tkod-Pol in PCR, suggesting further improvement within the Pfu-Pol framework is attainable. The significance of this work is the observation that improvements in PCR performance are easily attainable by blending elements from closely related archaeal polymerases, an approach that may, in future, be extended by using more polymerases from these organisms.

Keywords: PCR, *Pyrococcus furiosus*, *Thermococcus kodakarensis*, archaeal DNA polymerase, domain swapping, thermostability, fidelity

INTRODUCTION

Thermostable DNA polymerases are essential components of the polymerase chain reaction (PCR), a technique with myriad uses across the life sciences (Mullis et al., 1994; Weissensteiner et al., 2004; McPherson and Möller, 2006; Saunders and Lee, 2013). The family-B DNA polymerases from the *Thermococcales* order of the archaea are especially favored for PCR, due to their extreme stability at elevated temperatures and the presence of fidelity-conferring 3'-5' exonuclease activity (Lundberg et al., 1991; Cline et al., 1996; Takagi et al., 1997; Nishioka et al., 2001). Both the amino acid sequences and X-ray structures of these polymerases demonstrate a high degree of similarity, with structures available for the enzymes isolated from *Thermococcus gorgonarius* (Tgo-Pol) (Hopfner et al., 1999; Firbank et al., 2008; Killelea et al., 2010), *Thermococcus kodakarensis* (Tkod-Pol) (Hashimoto et al., 2001; Kuroita et al., 2005; Bergen et al., 2013), *Thermococcus* species 9°N-7 (9°N-Pol) (Chapin-Rodriguez et al., 2000), *Pyrococcus furiosus* (Pfu-Pol) (Kim et al., 2008) and

Pyrococcus abyssi (Pab-Pol) (Gougel et al., 2012). Despite similarities of amino acid sequence and structure, these polymerases have diverse kinetic properties; a strong influence on PCR performance. Most notably Tkod-Pol possesses higher processivity (the number of dNTPs incorporated per binding event) than other enzymes (Takagi et al., 1997; Nishioka et al., 2001). Enhanced processivity may arise from the presence of seven arginines, suggested to play a role in stabilizing primer-template binding and influencing the movement of DNA between the polymerization and proof reading active sites (Hashimoto et al., 2001; Kim et al., 2008). These arginines cluster near the "forked-point" (the junction between the template-binding and editing clefts) of Tkod-Pol (**Figure 1A**). The seven amino acids are well conserved in *Thermococcales* DNA polymerases, with two (R243 and R264) being present in all species (**Figure 1B**). The other five locations show more variation, although the amino acid at 266 is an arginine in both Tkod-Pol and Pfu-Pol. At the remaining four positions (247, 365, 381, 501), the arginines present in Tkod-Pol



are replaced by an alternative amino acid. The situation at Tkod-Pol position 381 is slightly obscured by insertion of an additional leucine at position 381 in Pfu-Pol. This arrangement, seen with many of the *Thermococcales* DNA polymerases, leads to two

possible sequence alignments (Figure 1B). The first, generated by the alignment algorithm Clustal (Sievers et al., 2011), lines up Tkod-Pol R381 with Pfu-Pol R382. However, superimposition of the structures of the two polymerases (Figure 1C) shows

that R381 in Tkod-Pol and L381 in Pfu-Pol are spatially equivalent, suggesting the alternative alignment shown in **Figure 1B** (Hashimoto et al., 2001). Another region that may contribute to processivity is the thumb domain, responsible for binding double-stranded DNA (Firbank et al., 2008; Killelea et al., 2010; Gougel et al., 2012; Bergen et al., 2013). This domain grips DNA tightly and interacts with newly synthesized double-strands, implying an important role in DNA translocation. In the absence of DNA, the thumb shows high flexibility and stretches of amino acids are often invisible in apo-enzyme crystal structures. The overall fold of the thumb domain does not differ significantly between archaeal DNA polymerases; however, its location relative to other domains is quite variable and entire domain motion is observed on DNA binding. Although the amino acid sequences in this domain are similar overall, many changes to individual amino acids are seen when Pfu-Pol and Tkod-Pol are compared (**Figure 1D**; supplementary data, Figure S1). However, in the current absence of polymerase-DNA-dNTP ternary structures, it is not easy to correlate any differences in processivity between Pfu-Pol and Tkod-Pol with individual amino acids in the thumb region.

During archaeal replication processivity is facilitated by PCNA, a sliding clamp which completely encircles the DNA, while simultaneously binding to the polymerase (Ishino and Ishino, 2011). Polymerase dissociation from DNA is hindered and overall processivity is enhanced, leading to rapid copying of the genome. In the *thermococcales* PCNA is a homo-trimeric ring-shaped molecule and the ring must be opened to load the protein onto DNA. *In vivo*, loading of PCNA is carried out by RFC, an ATP-driven clamp loader (Ishino and Ishino, 2011). However, *in vitro*, e.g., during the PCR, PCNA is not usually present and copying of DNA relies on the intrinsic processivity of the polymerase. Occasionally PCNA has been added and has been reported to enhance the PCR capability of Tkod-Pol (Kitabayashi et al., 2002). A recent comprehensive study showed that native Pfu-PCNA only improved PCR if RFC was also present, presumably to facilitate loading of the clamp onto DNA (Ishino et al., 2012). However, two PCNA mutants with reduced ring stability were able to self-load and considerably improved the PCR performance of Pfu-Pol, in the absence of RFC. For biotechnology applications more processive polymerases have been generated by fusion with double-stranded DNA binding proteins such as helix-hairpin-helix motifs (Pavlov et al., 2002) and Sso7d (a thermostable protein from *Sulfolobus solfataricus*) (Wang et al., 2004); the latter giving significant improvement in many PCR applications. Elevated processivity leads to more rapid copying of DNA and a polymerase with this property should be better at the PCR i.e., generate product more rapidly or produce longer amplicons. As processivity, *in vivo*, is largely a function of the sliding clamp, there may be little evolutionary pressure on polymerases to maximize this function, allowing considerable scope for improvement. In this publication we have sought to rationalize the differences in processivity and PCR ability previously observed between Tkod-Pol and Pfu-Pol. Hybrid polymerases have been produced, mainly Pfu-Pol but containing amino acids/domains from Tkod-Pol suggested to contribute to processivity. Do these hybrids result in a transfer of processivity from Tkod-Pol to Pfu-Pol and improved

PCR performance? A better comprehension of why two structurally similar DNA polymerases have different kinetic properties is intrinsically important and may also guide further rational mutagenesis.

MATERIALS AND METHODS

PROTEIN PURIFICATION AND MUTAGENESIS

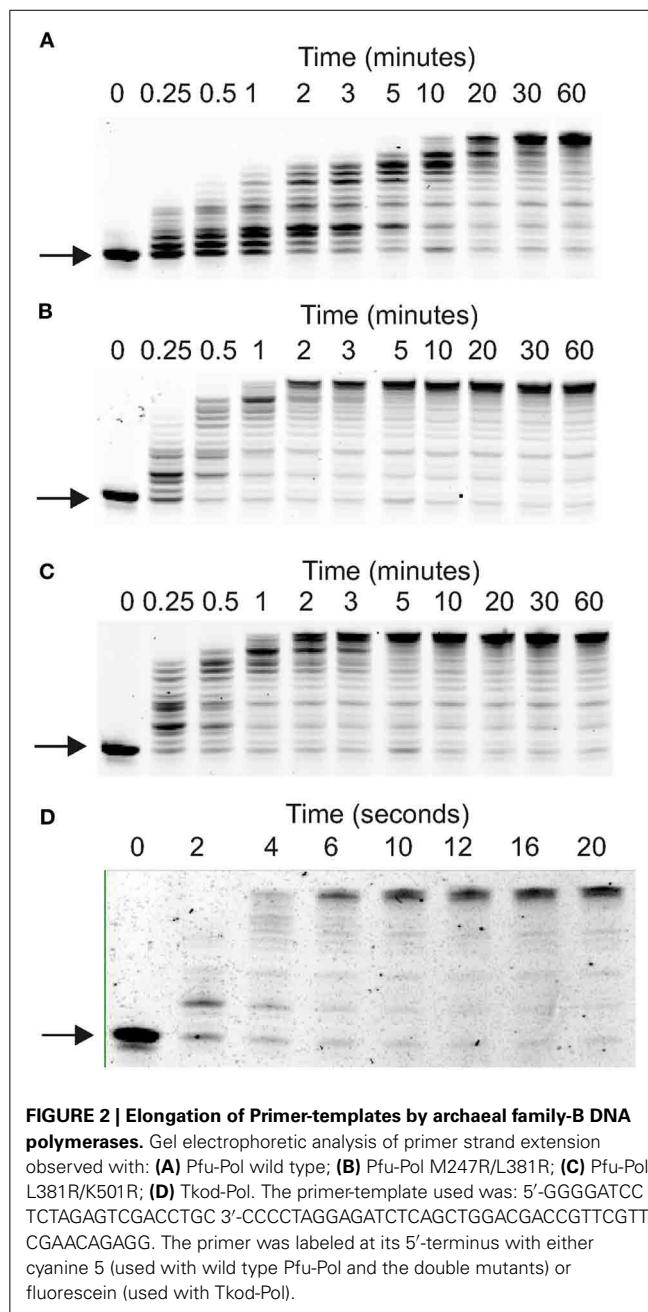
Wild type Pfu-Pol B (gene inserted into pET17b) was purified as reported (Evans et al., 2000; Emptage et al., 2008). Wild type TKod-Pol B (present in pET21a, supplied by Dr. Zvi Kelman University of Maryland) was purified in an identical manner. The single, double and triple mutants of Pfu-Pol were produced using a QuickChange® site directed mutagenesis kit (Agilent-Stratagene, Stockport) with Velocity™ DNA polymerase (Bioline, London, UK). The genes encoding the Pfu-TKod thumb swap derivatives were prepared using overlap extension PCR (Warrens et al., 1997). All mutated genes were completely sequenced to ensure the presence of the desired mutation and an absence of changes elsewhere. All mutants were purified in the same manner as the wild type enzyme with SDS-PAGE and Coomassie Blue staining indicating a purity of >95%.

POLYMERASE AND EXONUCLEASE ASSAYS

Primer-template extensions used the fluorescent-labeled primer-templates given in the legends to **Figures 2, 3** and supplementary data Figures S2, S3. These reactions were conducted, at 30°C, in 400 μl of 20 mM Tris pH 8.0, 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1% (v/v) Triton X-100 and 40 μg of bovine serum albumin. 400 μM of each dNTP was used with 10 nM primer-template and reactions were initiated by adding the polymerase to a final concentration of 500 nM. Subsequently 40 μl aliquots were sampled at appropriate time-points by quenching with an equal volume of stop buffer (95% formamide, 10 mM EDTA, 10 mM NaOH, 2 μM of a competitor oligodeoxynucleotide and 0.05% xylene cyanol indicator dye. Primer-templates were denatured by heating at 100°C for 10 min followed by cooling on ice. The competitor, which has the same sequence as the fully extended primer but lacks the fluorophore, prevents significant re-hybridization of the fluorescent products to the template (Russell et al., 2009). The samples (20 μl) were resolved on a denaturing 17% polyacrylamide gel containing 8 M urea run at 4.5 Watts for 4.5 h and visualized using a Typhoon scanner with ImageQuant software (GE Healthcare). Exonuclease assays were performed similarly, save for omission of the dNTPs and data were fitted to a first order reaction (% substrate remaining = 100e^{-kt} + offset; k = rate constant, t = time) using GraFit (Erithacus Software, London, UK), giving rate constants.

REAL TIME PCR

Real-time PCR used a Rotor-Gene 6000 thermocycler (Corbett Research, Qiagen). Genomic DNA from *S. cerevisiae* was used as the template, with the DNA Pol 2 gene targeted for amplification. A common forward primer (TACGTACCGCCGCAATACAA TGGCAGG) and four different reverse primers (TCGAATTG CCGCCGCCATTACTACCAC, TCGACTTGAAGCTCCCACC CTCTTCATC, GGCGTCAACTTTCCGAGGCCATTGC and TCATCGAACATGTCCAAGCCGTGAATCTTAC) were used to



amplify lengths of 145, 232, 543, and 1040 base pairs. Reactions were carried out in 25 μ l that contained 30 ng *S. cerevesiae* genomic DNA (Novagen), 1 μ M of each primer, 400 μ M of each dNTP, and 2.5 μ l of SYBR green (10,000 \times stock in dimethylsulfoxide, Invitrogen; initially diluted 1000-fold with water). The reactions were initiated by adding the polymerase (final concentration 20 nM) in the same buffer used for primer-template extensions. The PCR consisted of 1 \times 95°C for 2 min followed by 40 cycles of: 95°C (10 s)/58°C (20 s)/72°C (the time used for the 72°C extension step varied as indicated in the results). On completion of the PCR a melt curve analysis, consisting of a 90 s pre-melt step at 67°C followed by a temperature increase to 95°C at 0.2°C per second, was carried out. 20 μ l of the real

time PCR mixtures were run on a 1% agarose gel (detection with ethidium bromide) in order to verify that amplification resulted in a product of the correct size.

PCR

Amplification of a stretch of DNA ~5 kbases long within the plasmid pET17b[Pfu-Pol] (Evans et al., 2000) was carried out in 50 μ l using TCTGCTATGTGGCGCGGTATTATCC and CAACTCAGCTTCTTCGGGCTTG (1 μ M each) as primers, 400 μ M of each of the 4 dNTPs and either 20 or 100 nM of polymerase. 50 ng of pET17b[Pfu-Pol] was used with two buffers, either 20 mM Tris-HCl pH 8 or 20 mM Bicine-NaOH pH 9 both containing 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1% (v/v) Triton-X100 and 5 μ g bovine serum albumin. The PCR cycle comprised: pre-heat at 98°C (2 min); 30 cycles of 95°C (30 s), 60°C (30 s), 70°C (5 min); final hold at 70°C for 5 min. 20 μ l of the PCR mixtures (expected size of the correct amplicon ~5 kb) were analyzed using 1% agarose gel electrophoresis with ethidium bromide staining.

POLYMERASE FIDELITY

Fidelities were determined using pSJ2, an assay based on gap-filling of the single stranded *lacZ α* segment within the plasmid by a polymerase *in vitro*. The gapped derivative of pSJ2 (1 nM) was fully extended using the polymerase (100 nM) in 20 μ l of 20 mM Tris (pH 8.0), 10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgCl₂, 0.1% (v/v) Triton X-100, 2 μ g bovine serum albumin and 250 μ M of each of the four dNTPs. Extension was carried out at 70°C for 30 min and the mixture was used to transform *E. coli* Top 10 cells, which were plated on LB agar (containing X-gal, IPTG and ampicillin) and scored for blue/white colonies. Ratios of blue/white colonies were converted to mutation frequency and error rate as previously described (Jozwiakowski and Connolly, 2009; Keith et al., 2013).

PRIMER-TEMPLATE BINDING

The binding of the polymerases to primer-templates was determined using fluorescence anisotropy with hexachlorofluorescein-labeled DNA (Shuttleworth et al., 2004). The buffer conditions are detailed in the supplementary data (Figure S6).

PROCESSIVITY

Four hundred μ l of 20 mM Tris pH 8.0, 10 mM KCl, 10 mM (NH₄)₂SO₄, 1 mM EDTA, 0.1% Triton X-100, 40 μ g of bovine serum albumin, 400 μ M of each of the four dNTPs, 40 nM primer-template (sequence given in the legend to Figure 7) and 500 nM of polymerase were pre-incubated at 50°C for 5 min. Reactions were initiated by the simultaneous addition of 3 mM MgSO₄ and 10 μ M of a uracil-rich oligodeoxynucleotide trap (sequence given in legend to Figure 7) and polymerization allowed to proceed at 50°C. Aliquots of 40 μ l were withdrawn at appropriate times and extension determined using denaturing gel electrophoresis as described above.

DSF ANALYSIS

The thermal stabilities of the polymerases were analyzed by heating samples in a Rotor-Gene 6000 (Corbett Research, Qiagen) in the presence of SYPRO Orange and measuring any increase

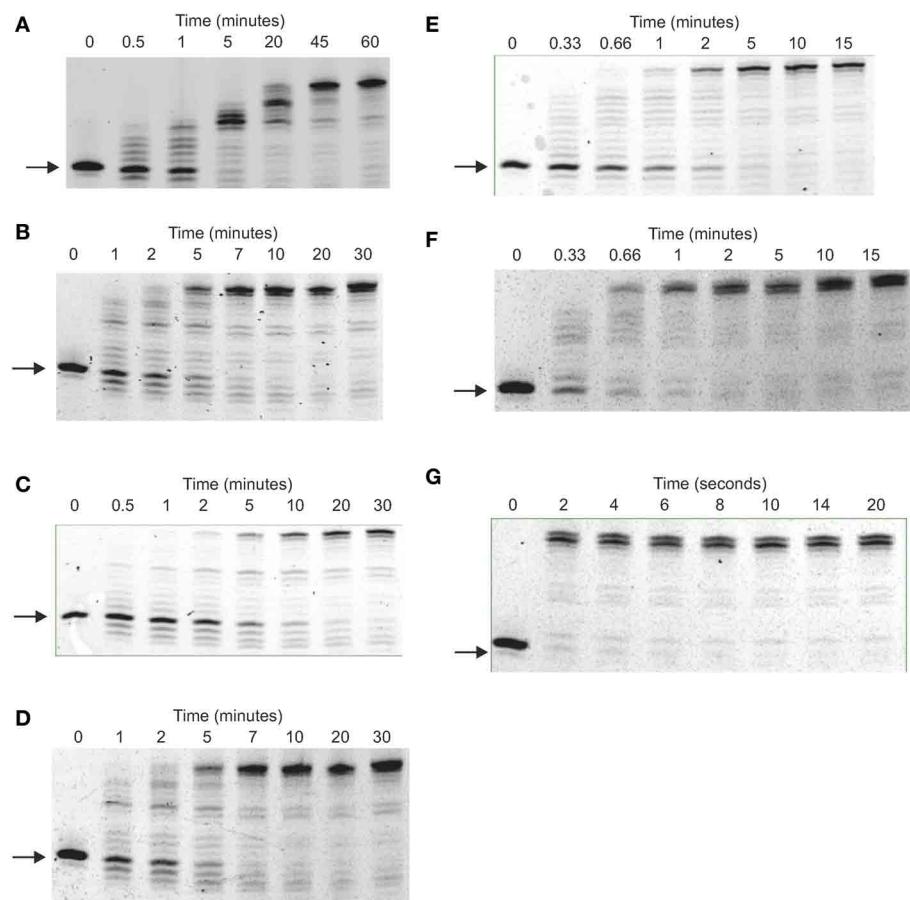


FIGURE 3 | Elongation of primer-templates by archaeal family-B DNA polymerases. Gel electrophoretic analysis of primer strand extension observed with: (A) Pfu-Pol wild type; (B) Pfu-Pol M247/L381; (C) Pfu-Pol L381R/K501R; (D) Pfu-Pol M247R/L381R/K501R; (E) Pfu-TkodTS;

(F) Pfu-TkodTS L381R/K501R; (G) Tkod-Pol wild type. The primer-template used was: 5'-GGGGATCCCTAGAGTCGACCTGCAGGGCAA 3'-CCCCTAGGA GATCTCAGCTGGACGTCCCGTTCGTTCAACAGAGG. The primer was labeled at its 5'-terminus with fluorescein.

in fluorescence. Thermal melts were carried out in 100 μ l of 40 mM HEPES, pH 7, containing 400 mM NaCl, 2 M guanidinium hydrochloride, 10 μ l of SYPRO Orange (5000 \times stock in dimethylsulfoxide, Sigma; initially diluted 100-fold with water). The polymerases were used at a concentration of 2 μ M. Excitation and emission were at 470 nm and 555 nm, respectively. The temperature increased from 25 to 100°C at a rate of 1°C per min. Data analysis was carried out with the Rotor-Gene 6000 series software and the melting profiles are presented as first derivatives.

RESULTS

MANIPULATING AMINO ACIDS AND DOMAINS OF Pfu-Pol TO MAKE IT MORE SIMILAR TO Tkod-Pol

Tkod-Pol contains seven arginines at the forked-point, four of which are replaced in Pfu-Pol (Figures 1A,B). The absent arginines have been investigated by introducing them into Pfu-Pol, initially as the single amino acid substitutions, Pfu-PolM247R, T265R, and K502R. The situation at Tkod-Pol position 381 is more complex, due to the insertion of an additional leucine at position 381 in Pfu-Pol (Figure 1B). Therefore, two Pfu-Pol variants have been created, corresponding to the potential

sequence line ups shown in Figure 1B. The deletion of L381, to give Pfu-PolL381 Δ , removes the extra amino acid, which may bring R382 in Pfu-Pol into register with R381 in Tkod-Pol. Additionally the direct substitution mutation, Pfu-PolL381R, has been prepared, corresponding more closely to structural data (Figure 1C). Insertion of L381 also means that lysine 502 in Pfu-Pol corresponds to arginine 501 in Tkod-Pol. Several of the single amino acid modifications to Pfu-Pol showed improved ability to copy DNA (next section). To generate incremental increases in polymerization activity, two double mutants, Pfu-PolM247R/L381R and Pfu-PolL381R/K502R and a triple variant, Pfu-PolM247R/L381R/K502R were prepared.

A second Pfu-Pol mutant has the thumb domain (amino acids T591 to S775 at the carboxyl terminus) replaced with the corresponding region from Tkod-Pol (named Pfu-TkodTS, TS = thumb swap). There are multiple differences in the amino acid sequences that comprise the thumb domains of the two polymerases (Figure 1D; supplementary data figure S1). Of the amino acids in the Tkod thumb that contact DNA directly, only two, R709 and G711 (shown in green in Figure 1D), are changed in Pfu-Pol, to proline and serine, respectively. However, many Tkod

residues, immediately adjacent to a DNA-contacting amino acid, also vary in Pfu-Pol (supplementary data Figure S1). The numerous variations in sequence make it a sizeable task to probe the contribution of individual amino acids; rather a complete thumb transplant was used. Finally the most advantageous “forked-point” double mutant (L381R/K502R) has been combined with the thumb domain swap.

PRIMER-TEMPLATE EXTENSION BY DNA POLYMERASES

Polymerase activities were initially compared using several primer-templates, a simple extension assay useful for indicating the effectiveness with which DNA is copied. The single amino acid substitutions Pfu-Pol M247R and K502R generated full-length product more rapidly than wild type (supplementary data, Figure S2). At amino acid position 381, where two mutations were made, L381R appeared slightly better than L381Δ and both were superior to wild type. Pfu-Pol T265R was the only instance of a substitution giving poorer extension (supplementary data, Figure S3). Building on the single-swaps, two double mutants (M247R/L381R and L381R/K502R) were created; both extended the primer-template significantly more rapidly than wild type and slightly faster than their single parents. Faster extension with the double mutants was consistently observed with three primer-templates (**Figures 2, 3**; supplementary data Figure S2). As can be seen in **Figure 2**, wild type Pfu-Pol required 10–20 min before full length product became visible, whereas completely extended primer was apparent after 1–2 min with both of the doubles. An alternative primer-template (**Figure 3**) again indicated more rapid extension with the double mutants (compare the five and twenty minute time points for wild type, M247R/L381R and L381R/K502R in **Figure 3**). This figure additionally demonstrates that the triple mutation (M247R/L381R/K502R) shows little additive improvement. Although Pfu-Pol M247R/L381R and L381R/K502R are superior to wild type, they do not match the performance of Tkod-Pol, where full extension is apparent after less than 10 s (**Figures 2, 3**). The assay was also shown that Pfu-TkodTS extends primer-template more rapidly than wild type (**Figure 3**). With Pfu-TkodTS the reaction is almost complete after 5 min, whereas little full length product is seen with the wild type after this time. Even faster extension was seen when Pfu-TkodTS/L381R/K502R was investigated, with the reaction finishing in 1–2 min. Extensions were carried out at 30°C, away from the temperature optima of about 75°C for Pfu-Pol and Tkod-Pol (Takagi et al., 1997). Maybe, as temperature decreases, the activity of Pfu-Pol declines more steeply than Tkod-Pol, explaining the observed superior performance of Tkod-Pol at 30°C. However, Tkod-Pol maintained its advantage in both real time and standard PCR (**Figures 4, 5** respectively), techniques that involve DNA synthesis near the temperature optimum. Based on primer-template extensions, four mutants, M247R/L381R, L381R/K502R, Pfu-TkodTS, and Pfu-TkodTS/L381R/K502R were selected for more detailed investigation.

REAL TIME PCR ANALYSIS

A key motivation in manipulating Pfu-Pol was to produce superior PCR enzymes and, to allow direct identification of useful mutants, priority was given to PCR-based assays. Initially real

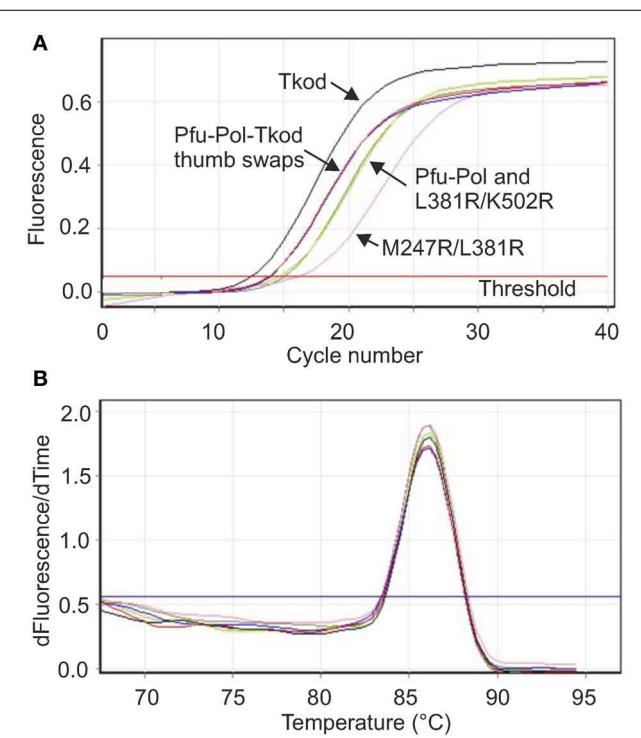
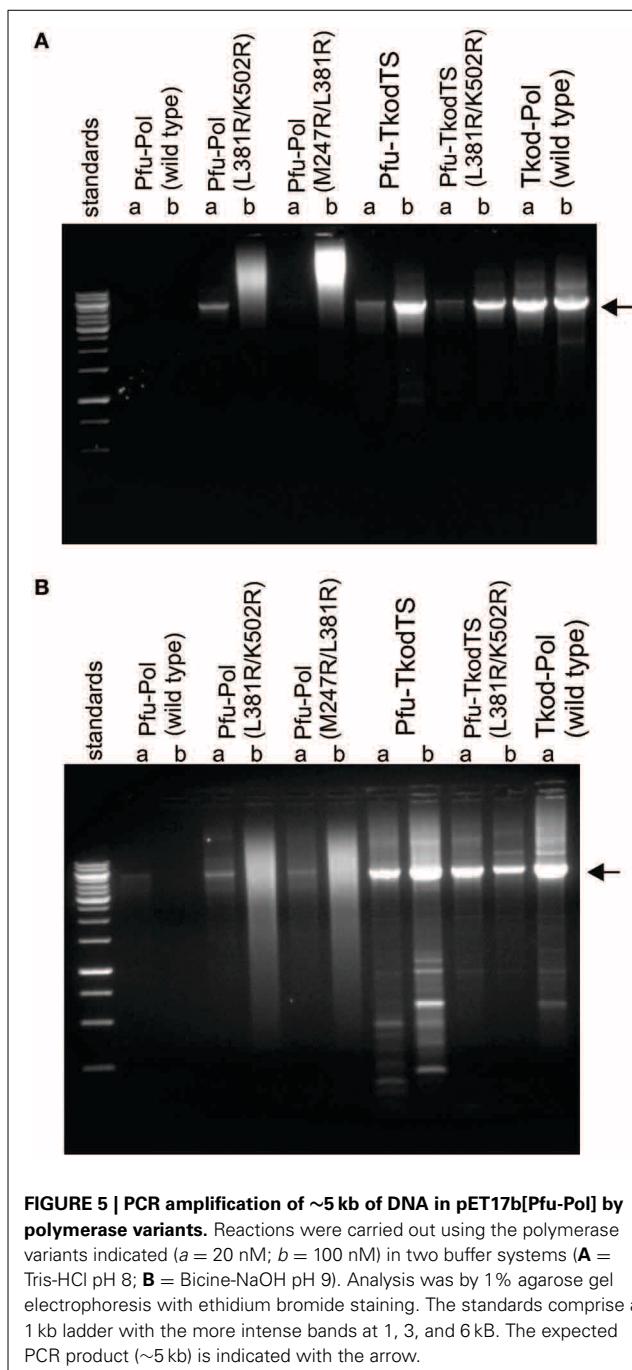


FIGURE 4 | Real time PCR analysis of polymerase performance.

(A) Amplification of a stretch of yeast genomic DNA 232 bases in length using 10 s extension. The lines that correspond to the individual polymerases are identified on the figure. Pfu-Pol TS and Pfu-Pol TS L381R/K502R (indicated Pfu-Tkod thumb swaps) gave near superimposable lines. Likewise the lines for Pfu-Pol and the double mutant L381R/K502R overlapped strongly. (B) Melting temperature analysis (first derivative showing the rate of change of temperature with time against time) of the amplicons generated in (A). All the polymerases gave exclusively the desired product as indicated by a single peak with a T_m of 86°C. As all the lines are essentially identical the individual polymerases have not been identified. In both panels only a single line for each polymerase is shown but all experiments were carried out in triplicate.

time PCR (RT-PCR), commonly used to measure DNA and RNA levels (Saunders and Lee, 2013), was applied to investigate polymerase performance. During RT-PCR a C_t value defines the number of cycles taken for product to become apparent; the more effective the polymerase, therefore, the lower the C_t . In these experiments, yeast genome DNA targets of 232, 543, and 1040 bases were generated and the products detected using the fluorophore SYBR green. With the shortest amplicon (232 bases), Pfu-Pol and the double mutant L381R/K502R generated product after a similar number of cycles (C_t 15.30 and 14.93, respectively) and M247R/L381R was slightly slower (C_t 15.96). Here, the behavior of M247R/L381R is anomalous as it appeared superior to the wild type in every other test of polymerization ability. Tkod-Pol produced product more rapidly than the wild type i.e., had a lower C_t , and the two thumb swap mutants showed intermediate C_t values. As an example the real time PCR data found with the short amplicon is shown in **Figure 4** and all C_t values are summarized in **Table 1**. Further information was obtained using a longer amplicon of 543 bases and varying the extension time.



The double mutants appeared superior to Pfu-Pol, either giving a product where none was produced by the parent or generating amplicons more quickly; L381R/K502R demonstrated better performance than M247A/L38R. Greater efficiency was again apparent for Tkod-Pol, which generated products more rapidly as measured by C_t values. Similarly the thumb swap mutants showed better performance than Pfu-Pol (wild type and the double mutants) but were inferior to Tkod-Pol (Table 1). Confirmation of these trends came from the use of the longest amplicon (1040 bases) with four different extension times. Overall, as summarized in Table 1, Tkod-Pol gives the best performance in RT-PCR

followed, in order, by Pfu-TkodTSL381R/K502R, Pfu-TkodTS, Pfu-PolL381R/K502R, Pfu-PolM247R/L381R, with Pfu-Pol wild type, the worst enzyme. This ranking is an excellent match to the results seen in primer-template extensions. While RT-PCR is a straightforward method for characterizing polymerases, incorrect products are often produced. Therefore, the integrities of the 232, 543, and 1040 base products were checked using melting temperature analysis (all three had similar T_m values of between 86 and 88°C; for an example see Figure 4) and by gel electrophoresis to confirm the presence of product with the expected length (supplementary data, Figure S4). In Table 1, C_t values are only quote if the anticipated product comprised at least 95% of the amplification mixture.

PCR PERFORMANCE OF POLYMERASES

To determine if the more rapid primer-template extensions and lower C_t values exhibited by the Pfu-Pol mutants translated into improved standard PCR performance, amplifications of a 5 kb stretch of DNA were carried out. Two buffers with pH of 8 and 9 were used and the polymerases were tested at two concentrations, 20 or 100 nM. The results are given in Figure 5 and show that wild type Pfu-Pol failed to generate substantial amounts of product, traces being apparent only at pH 9 with 20 nM enzyme. The double mutants M247A/L381R were better; at 20 nM levels both gave obvious product at pH 9, L381R/K502R also yielded product at pH 8. However, at the higher concentration of 100 nM the doubles gave smeared bands, suggesting non-specific amplification. The thumb swaps showed marked PCR improvement, at both pH values and protein concentrations a clear and intense (apart from pH 8, 20 nM) product band was visible, suggesting efficient amplification. Some non-specific products were apparent, particularly at pH 9. Wild type Tkod-Pol also performed well, perhaps marginally better than the two thumb swap mutants. Non-specific smaller products were also observed with Tkod-Pol at pH 9. The PCR results concur with all those above; the double mutants are superior to wild type Pfu-Pol and the thumb swaps even better. In these experiments the thumb swaps approach the performance exhibited by wild type Tkod-Pol.

FIDELITY OF DNA POLYMERASES

DNA polymerase accuracy is crucial for PCR and the main reason archaeal enzymes are widely applied. Two techniques were used to check the fidelity of the enzymes. Initially proof reading exonuclease activity was directly measured using a primer-template containing a base mismatch at the point of extension. A protein concentration (500 nM) in excess of primer-template (10 nM) was used and exonuclease activity observed by monitoring the removal of the mismatched base at the 3'-end of the primer. All Pfu-Pol mutants demonstrated slightly faster exonuclease rate constants than wild type, although increases were at most a factor of two (Figure 6, Table 2). Noticeably wild type Tkod-Pol showed more pronounced activity, about 7.5 times faster than Pfu-Pol. The plots shown in Figure 6 and the rate constants in Table 2 measure the disappearance of initial primer and so report on removal of the mismatched base. It is abundantly clear that subsequent degradation of the primer can occur, as a consequence of exonuclease activity at Watson-Crick base pairs. With Tkod-Pol

Table 1 | RT-PCR performance of the polymerase variants.

Experimental conditions		C_t value of polymerase variant ^b					
Length of amplicon ^a	Extension time (s)	Pfu-Pol wild type	Pfu-Pol M247R/L381R	Pfu-Pol L381R/K502R	Pfu-TodTS	Pfu-TkodTS L381R/K502R	Tkod-Pol wild type
232	10	15.30	15.96	14.93	13.90	14.07	12.58
543	10	NP	NP	NP	14.65	12.73	10.94
543	30	NP	12.20	10.55	8.61	8.07	6.72
543	60	11.78	11.07	10.04	8.60	8.04	6.52
543	90	11.16	10.46	9.66	8.54	7.80	6.54
1040	10	NP ^c	NP	NP	NP	NP	12.29
1040	30	NP	NP	NP	NP	16.90	11.05
1040	60	NP	NP	16.37	14.51	12.74	10.83
1040	90	NP	NP	13.85	12.66	11.10	10.75

^aYeast genome DNA was the target for amplification and primers (given in Materials and Methods) were selected to give the amplicon lengths indicated.

^bThe figures in the table represent the C_t value, the number of cycles required for the amplicon to become detectable. All experiments were conducted in triplicate with the C_t being the average of the three runs. In all cases where a figure is quoted, melting temperature analysis revealed a product with a T_m value of between 86 and 88°C and gel electrophoresis showed an amplicon of the correct length (Figure 4; supplementary data, Figure S4). In all cases the anticipated product comprised at least 95% of the total amplified material.

^cNP, no product; either no product was produced or non-specific amplification occurred, giving either an incorrect product or a mixture of amplicons containing both the desired and non-specific products.

accumulation of the n-1 product is clear, suggesting preferential degradation of mismatched bases. Equivalent accumulation is less obvious with Pfu-Pol, inferring less discrimination between mismatches and *bona fide* base-pairs in this case. To further investigate the proof reading activities of Pfu-Pol and Tkod-Pol were compared using a fully complementary primer-template. With these undamaged substrates the exonuclease activity of Tkod-Pol was only 1.5 fold faster than Pfu-Pol, less than the factor observed with the mismatched substrate (supplementary data, Figure S5). The faster exonuclease rate of Tkod-Pol, compared to Pfu-Pol, with mismatches, followed by near equal reactivity at normal bases accounts for the n-1 product accumulation seen in Figure 6. Exonuclease activities were greater with mismatched than fully complementary substrates (Table 1); an expected observation as mismatches are easier to unwind, a necessary step for proof reading (Reha-Krantz, 2010).

To complement the measurements of exonuclease activities, fidelities have been determined using a *lacZα* plasmid-based assay (Jozwiakowski and Connolly, 2009; Keith et al., 2013). A gapped plasmid, which contains the *lacZα* gene in the single stranded region, is fully extended, *in vitro* at 70°C, by a polymerase and then used for transformation of an appropriate *E. coli* strain. The observed ratio of white to blue colonies on plates contain the indicator X-gal, reflects the fidelity of the polymerase. For these experiments pSJ2 was used, a plasmid that has been characterized sufficiently to allow conversion of white/blue ratio i.e., the mutation frequency into error rate, the number of mistakes made per base incorporated (Keith et al., 2013). The results observed are given in Table 3, which demonstrates near identical error rates for each of the polymerases. All the mutants maintain the high accuracy associated with wild type Pfu-Pol and Tkod-Pol. Although Tkod-Pol has a measurably higher exonuclease activity than all the Pfu-Pol derivatives (Table 2), this does not translate into higher accuracy as determined using pSJ2. Fidelity critically depends on the balance between exonuclease and polymerase

rates and with Tkod-Pol both are elevated compared with Pfu-Pol. The ratio of polymerase/exonuclease activities may be similar for both enzymes, leading to equal propensity to continue synthesis vs. engaging proof-reading and conferring similar overall fidelity on both proteins.

PRIMER-TEMPLATE BINDING

The influence of the mutations on primer-template binding was determined using fluorescence anisotropy with hexachlorofluorescein-labeled primer-templates. Previous studies, carried out at pH 7.5 and 100 mM NaCl (ionic strength = 100), indicated relatively poor binding of Pfu-Pol to primer-templates, a K_D of 270 nM being measured (Shuttleworth et al., 2004). As expected, binding affinity was increased at the lower salt concentration of 20 mM KCl (ionic strength = 20), reflected in a K_D of 32 nM (Richardson et al., 2013). In this publication binding titrations were performed in the same buffer used for extensions; pH 8.5 containing 10 mM KCl and 10 mM $(\text{NH}_4)_2\text{SO}_4$ (ionic strength of 40). Under these conditions wild type Pfu-Pol bound the primer-template with a K_D of 251 nM (supplementary data, Figure S6). Tkod-Pol showed similar affinity (K_D = 276 nM) and so the more rapid primer-template extensions seen with this enzyme, cannot simply be accounted for by tighter binding to the DNA substrate. The two double mutants showed better interaction with DNA, by a factor a little greater than two for L381R/K502R and just over three for M247R/L381R (supplementary data, Figure S6). Surprisingly Pfu-KodTS bound primer-template (K_D = 95 nM) more strongly than either of the parent polymerases from which it is derived and a further improvement was observed with the Pfu-KodTS L381R/K502R (K_D = 50 nM). All The K_D values seen are summarized in Table 2.

PROCESSIVITY OF DNA POLYMERASES

The processivity of a polymerase is the number of dNTPs incorporated per binding event (Von Hippel et al., 1994; Bambara et al.,

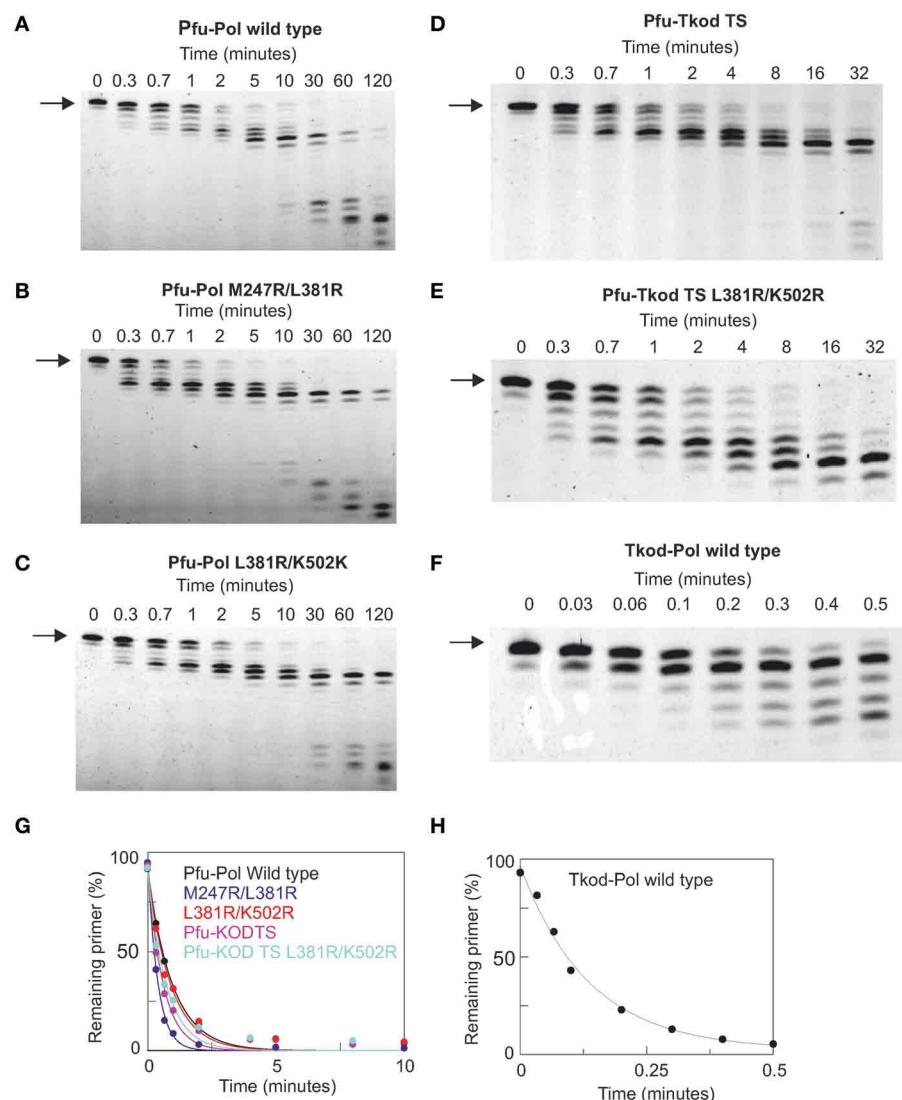


FIGURE 6 | Proof reading 3'-5' exonuclease activity by archaeal family-B DNA polymerases. (A–F) Gel electrophoresis analysis of primer strand degradation seen with a base mismatch primer-template using the polymerase variants indicated. The starting primer is shown arrowed. Panels (G) and (H) show fits of the data (as % primer remaining with time) to a single

exponential decay to give the rate constants summarized in **Table 2**. The primer-template used was: 5'-GGGGATCCTCTAGAGTCGACCTGC 3'-CCCCTAGGAGATCTAGCTGGACACCGTTCGAACAGAGG. The primer was labeled at its 5'-terminus with fluorescein. The mismatched bases are shown underlined,

1995). For accurate measurement each primer should undergo only one round of extension and, under these single hit conditions, the number of dNTPs incorporated equals the processivity. Single hits can be achieved using a low concentration of polymerase, relative to primer-template, such that the probability of secondary initiation is low. Alternatively a trap, such as heparin or poly(dA-dT), can be used to sequester the polymerase after it dissociates from the primer-template (Bambara et al., 1995). Archaeal DNA polymerases bind tightly to uracil-containing DNA (Shuttleworth et al., 2004), enabling a 23-mer containing 5 uracil residues (**Figure 7**) to be used as trap. The uracil-rich trap bound all polymerase variants with K_D values in the 1–10 nM range (data not shown). For processivity

determination 40 nM primer-template (sequence given in the legend to **Figure 7**) and 500 nM polymerase were pre-incubated in the absence of Mg^{2+} and reactions initiated by the simultaneous addition of the metal and 10 μ M of the trap. These stringent conditions, a 250-fold excess of a uracil-rich oligodeoxynucleotide with a high affinity for polymerase, should ensure efficient polymerase sequestration and inhibit re-binding to primer-template.

All polymerases were observed to have very low processivities at 250-fold trap excess and nearly identical results were seen with a 100-fold excess (data not shown). Pfu-Pol acted largely in a distributive manner, with little product seen at short time intervals (**Figure 7A**; larger versions of all processivity gels are given in

Table 2 | Summary of the properties of the polymerases investigated.

DNA-Pol	K_D (nM) ^a	k_{exo} (min ⁻¹) (mis-paired DNA) ^b	k_{exo} (min ⁻¹) (base-paired DNA) ^c	Processivity ^d	Melting temperature (°C) ^e
Pfu-Pol wild type	251 ± 23	1.0 ± 0.1	0.17 ± 0.02	1	93.2 ± 0.5
Pfu-Pol M247R/L381R	70 ± 9	2.2 ± 0.1	nd	1	95.7 ± 0.2
Pfu-Pol L381R/K502R	107 ± 13	1.2 ± 0.1	nd	1–3	94.5 ± 0.8
Pfu-TkodTS	95 ± 7	1.8 ± 0.1	nd	1–3	95.3 ± 1.0
Pfu-TkodTS L381R/K502R	50 ± 7	1.7 ± 0.1	nd	3	95.3 ± 0.6
Tkod-Pol wild type	276 ± 18	7.6 ± 0.2	0.24 ± 0.01	3	82.9 ± 0.3

^aThe affinity for DNA measured using fluorescence anisotropy titrations (supplementary data, Figure S6). Each value is the average of three determinations ± the standard deviation. The primer-template used had the following sequence (Hex = hexachlorofluorescein): 5'-HexGGGGATCCTCTAGAGTCGACCTGC 3'-CCCTCTAGGAGATCTCAGCTGGACGACCGTTCGTCAACAGAGG. (the mismatched bases are underlined.)

^bRate observed for the degradation of a mis-paired primer-template, determined with polymerase in excess of DNA (Figure 6). Each value is the average of three determinations ± the standard deviation. For the mis-paired DNA, the same sequence as given above was used with the underlined dG changed to dA, giving a dC:dA mis-match at the primer-template junction. Fluorescein (Flu) was used as indicator dye.

^cRate observed for the degradation of a fully base-paired primer-template, determined with polymerase in excess of DNA (supplementary data, Figure S5). Each value is the average of three determinations ± the standard deviation. This primer-template has the exact sequence given above and fluorescein (Flu) was used as indicator dye.

^dProcessivity (number of dNTPs incorporated per binding event) of the polymerases measured using a uracil-containing single stranded DNA trap (Figure 7).

^eMelting temperatures were determined by DSF in the presence of 2M guanidinium hydrochloride (Figure 8). The T_m of the first transition observed is given as an average ± standard deviation from three measurements.

Table 3 | Fidelities of DNA polymerases determined using the *lacZα* indicator pSJ2.

DNA-Pol	Total colonies ^a	Mutant (white) colonies ^b	Mutation frequency	Error rate ^c
			(corrected) ^b	
Pfu-Pol wild type ^d	25,700	11	3.2×10^{-4}	1.6×10^{-6}
Pfu-Pol M247R/L381R	20,555	8	2.8×10^{-4}	1.4×10^{-6}
Pfu-Pol L381R/K502R	26,675	9	2.3×10^{-4}	1.2×10^{-6}
Pfu-TkodTS	39,814	16	2.9×10^{-4}	1.5×10^{-6}
Pfu-TkodTS L381R/K502R	31,704	12	2.7×10^{-4}	1.3×10^{-6}
Tkod-Pol wild type	28,028	11	2.8×10^{-4}	1.4×10^{-6}

^aSum of three independent experiments, each consisting of five repeats.

^bThe Mutation frequency is the ratio mutant (white) colonies/total colonies and has been corrected by subtracting the background mutation frequency of 1.1×10^{-4} found for gapped pSJ2 (Keith et al., 2013).

^cThe Error rate is the number of mistakes made by the polymerase per base incorporated. The determination of the error rate from the mutation frequency has been described previously (Keith et al., 2013).

^dData taken from an earlier publication (Keith et al., 2013).

the supplementary data, Figure S7). A very faint band representing a single dNTP addition is seen after 2 min and +1 and +2 bands are visible after 5 min. Longer products were observed at 15 and 30 min but may result from multiple binding events. Thus the uracil trap is not perfect and at longer times polymerase “escape” is seen, maybe arising from trap degradation by the exonuclease activity. As expected, polymerization was more extensive when the trap was omitted (Figure 7A). Adding the trap in the pre-incubation step results in the abolition of the barely extended bands seen at short times; instead only slower mobility bands are observed after prolonged incubation, which correspond

well with the similar bands observed when the uracil trap is used in the standard manner. Pfu-Pol has a processivity of 1 (values are summarized in Table 2), implying dissociation prior to adding even a single dNTP is highly likely. With Tkod-Pol a prominent band, corresponding to the incorporation of 3 dNTPs is seen at the shortest time (5 s) in the presence of the trap, representing the processivity (Figure 7E, Table 2). Tkod-Pol is more active than Pfu-Pol and even in the presence of trap full length product, presumably due to multiple binding events, was seen at short times. As expected these bands were more prominent when the trap was left out. Pronounced exonuclease activity, evidenced by shortened primer products was also apparent. The four mutants showed processivity profiles near those of the wild types (Figures 7B–D, Table 2). The double mutant M247A/L381R was very similar to Pfu-Pol with a faint band, representing incorporation of a single dNTP just visible at the one minute time point. L381R/K502R appeared marginally better, the band indicating one dNTP addition was predominant at times of 1 and 2 min, but fainter bands corresponding to the +2 and +3 products were seen. The processivity of this mutant, 1–3, may be slightly improved. Processivity does not need to be a single integral value; rather polymerases show a spread of dNTP incorporations per binding event. Pfu-Kod(TS) also demonstrated a processivity of between 1 and 3 (Figure 7C). Finally Pfu-TkodTS L381R/K502R appeared to be equivalent to Tkod-Pol with a single band at the +3 location visible at short times. Overall Tkod-Pol has higher processivity than Pfu-Pol (3 vs. 1) and with the mutations values are shifted toward Tkod-Pol.

The processivities determined here, using a uracil-rich oligodeoxynucleotide trap, are much lower than previously observed. Values of 270 and >300 have been reported for Tkod-Pol and of 6, <20 and 80 for Pfu-Pol (Takagi et al., 1997; Wang et al., 2004; Kim et al., 2007). None of these studies used a trap to hinder re-binding of the polymerase and secondary extension

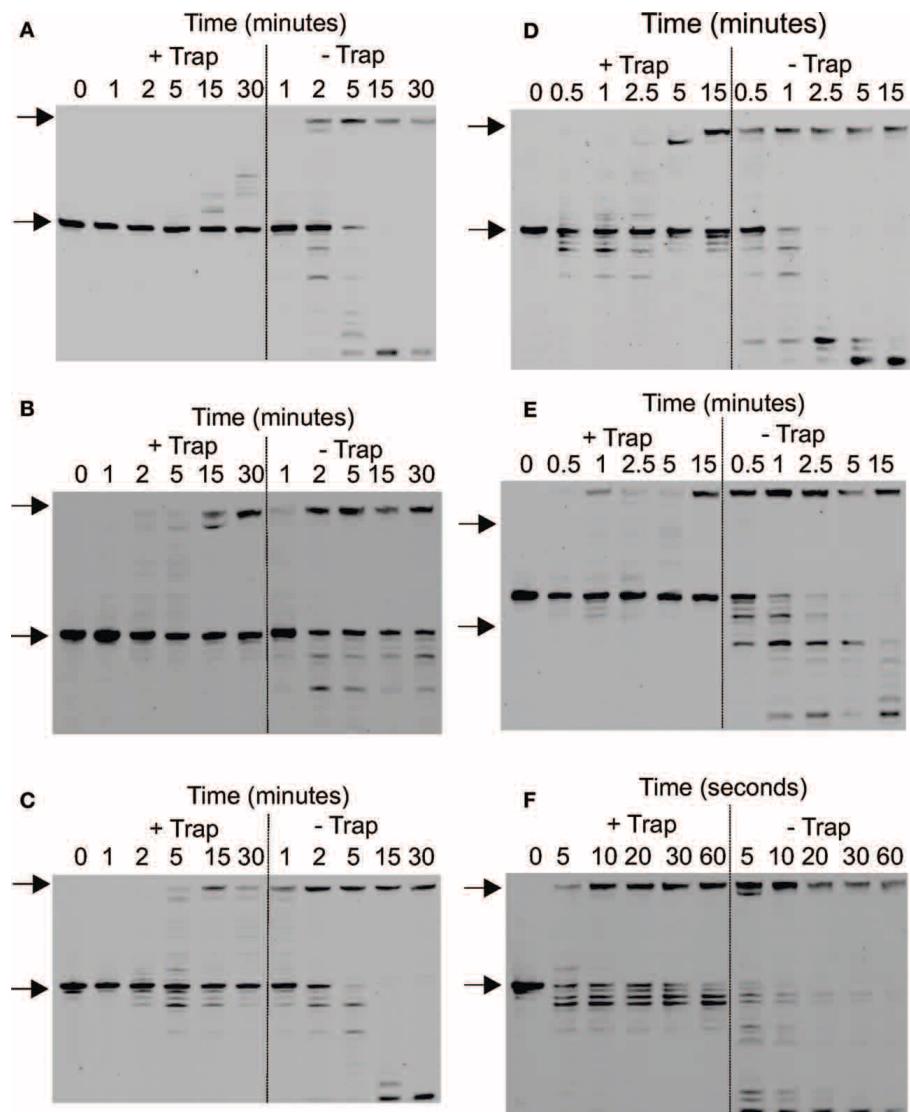


FIGURE 7 | Determination of the processivity of archaeal DNA polymerases. (A–F) Gel electrophoresis analysis of primer strand extension seen using the polymerase variants and times indicated for: **(A)** Pfu-Pol wild type; **(B)** Pfu-Pol M247/L381; **(C)** Pfu-Pol L381R/K501R; **(D)** Pfu-TkodTS; **(E)** Pfu-TkodTS L381R/K501R; **(F)** Tkod-Pol wild type. The starting primer and fully extended product are both arrowed. The primer-template used was: 5'-GGGGATCCTCTAGAGTCGACCTGC 3'-CCCTAGGAGATCTCAGCTGGACGA

CCGTTCGTTCGAACAGAGTACCTGGCTAT The primer was labeled at the 5'-terminus with fluorescein and the reaction initiated by the simultaneous addition of Mg²⁺ and the uracil rich single-stranded trapping oligodeoxynucleotide 5'-GTTGGUACUCTUAGUCTUTAGGT (extensions labeled + trap). For the extensions labeled—trap the competitor was omitted. Larger versions of each of the gel are given in the supplementary section (supplementary data Figure S7).

and the change in protocol may explain the discrepancy. All studies agree, however, that Tkod-Pol is more processive than Pfu-Pol. Yeast DNA polymerase δ, similar in structure to the archaeal enzymes (Swan et al., 2009), shows a processivity of 2–3 when evaluated using a trap (Hogg et al., 2014).

THERMOSTABILITY OF DNA POLYMERASES

The ability to survive elevated temperature is an essential polymerase feature for the PCR. Both archaea, from which the polymerases used in this study are isolated, are hyper-thermophiles, but their preferred growth temperatures differ. *Pyrococcus furiosus*

grows optimally at 100°C and can survive at 110°C, whereas *Thermococcus kodakarensis* grows best at 85°C and can tolerate 94°C (Fiala and Stetter, 1986; Borges et al., 2010). Our group earlier measured the thermostability of Pfu-Pol using differential scanning fluorimetry (DSF), where the protein is subject to a steady increase in temperature in the presence of the dye SYPRO orange (Killelea and Connolly, 2011). As thermally-induced unfolding takes place the dye binds to exposed hydrophobic regions, resulting in an increase in fluorescence that can be measured using a real time PCR apparatus. Pfu-Pol was found to be extremely thermostable with the unfolding transition being

incomplete at 100°C, the maximum achievable with this technique. Addition of guanidinium hydrochloride (GuHCl) destabilizes Pfu-Pol, making thermal unfolding more accessible. A Pfu-Pol mutant, that lacked two disulphide bridges and was more heat sensitive, demonstrated two well separated melting transitions. Most likely the wild type behaves similarly but the second transition cannot be observed as it takes place above 100°C (Killelea and Connolly, 2011). In the present study 2M GuHCl was used to bring heat-induced unfolding into the DSF range and the melting profiles are shown in **Figure 8**. Tkod-Pol shows two melting transitions, the first of which has a T_m of 82.9°C (all T_m values are summarized in **Table 2**). Pfu-Pol is more stable with a T_m of 93.2°C. As postulated earlier it is probable that a second melting event for Pfu-Pol takes place above 100°C; therefore, only the first observed transitions have been used for comparing unfolding. The T_m of 93.2°C is slightly higher than that of 89.7°C observed earlier for Pfu-Pol in 2 M GuHCl (Killelea and Connolly, 2011). Such variation may be accounted for by the different NaCl concentrations used, 400 mM here, 200 mM previously. Unsurprisingly the double mutants of Pfu-Pol, M247R/L381R ($T_m = 95.7^\circ\text{C}$) and L381R/K502R ($T_m = 94.5^\circ\text{C}$) retain the thermostability of the wild type, if anything being slightly more heat resistant. More unexpectedly the thumb swaps, in which about 24% of the protein is derived from Tkod-Pol, fully retain the stability of Pfu-Pol, with no lowering of T_m , toward that of Tkod-Pol is (**Figure 8** and **Table 2**).

DISCUSSION

The use of the PCR across biological, medical, veterinary, agricultural and forensic sciences has aroused considerable interest

in thermostable DNA polymerases and many enzymes, especially from the *Thermococcales* order of the archaea, are employed (Terpe, 2013). Tkod-Pol and Pfu-Pol are the most established and several investigations have pointed out the advantageous nature of the former in terms of speed and processivity (Takagi et al., 1997; Nishioka et al., 2001; Kim et al., 2007). The abundance of arginines at the “forked-point” has been offered as one reason for the high processivity of Tkod-Pol (Hashimoto et al., 2001; Kim et al., 2008), but the absence of a closed ternary complex (enzyme/primer-template/dNTP) limits knowledge on the exact functions of these amino acids. It was originally proposed that Tkod-Pol R247 may separate the primer-template and stabilize the denatured substrate, in a similar manner to corresponding amino acid in the bacteriophage RB69 polymerase (Shamoo and Steitz, 1999; Hashimoto et al., 2001). More recent studies have suggested that this amino acid is relatively unimportant (Aller et al., 2011; Richardson et al., 2013). The thumb domain of archaeal family-B polymerases, and indeed all DNA polymerases, is responsible for binding double-stranded DNA and expected to be critical in DNA translocation and processivity. Several “forked-point” arginines are missing in Pfu-Pol and the thumb domain shows subtle differences between the two enzymes. A major aim of this study was to graft these elements into Pfu-Pol in the hope of creating better PCR enzymes.

As “forked-point” arginines and/or the thumb domain of Tkod-Pol are introduced into Pfu-Pol, extension rates become faster, PCR performance improves and processivity, albeit only from one to three, is increased. In all cases the superior thermostability of Pfu-Pol is retained and there is no decrease in fidelity. The influence appears cumulative with the best mutant, combining both “forked-point” and thumb alterations. Pfu-TkodTS L381R/K502R represents an end-point (the entire thumb has been exchanged and addition of further missing arginines e.g., R246 does not enhance properties), yet still falls short of the performance of Tkod-Pol. This suggests other regions of the polymerase may play a role and these may subtly differ between Tkod-Pol and Pfu-Pol. One important area is the fingers domain, made up of two long α -helices (Hopfner et al., 1999; Chapin-Rodriguez et al., 2000; Hashimoto et al., 2001; Kim et al., 2008), which undergoes a conformational change following dNTP binding to the polymerase/primer-template complex, to produce a catalytically competent “closed” ternary complex. The fingers domain is highly conserved in all DNA polymerases and plays a critical role in dNTP selection and accurate insertion of the incoming base into the extending primer (Brautigam and Steitz, 1998). The amino acid sequences of the fingers domain of Tkod-Pol and Pfu-Pol are similar, but not identical (supplementary data, Figure S1). Given the critical function of this area in catalysis, the small variations may impinge on PCR performance. A second key element may be the Y-GG/A motif (Brautigam and Steitz, 1998), shown to be important in the processivity, fidelity and PCR capability of archaeal polymerases (Bohlke et al., 2000). The sequences near this region are shown in **Figure 1**, corresponding to the amino acids near arginine 381 (Y-GG/A = YEGG for Tkod-Pol and YTGG for Pfu-Pol). The presence of an additional leucine (L381) in Pfu-Pol alters the location of the conserved tyrosine and glycines, moving them away from

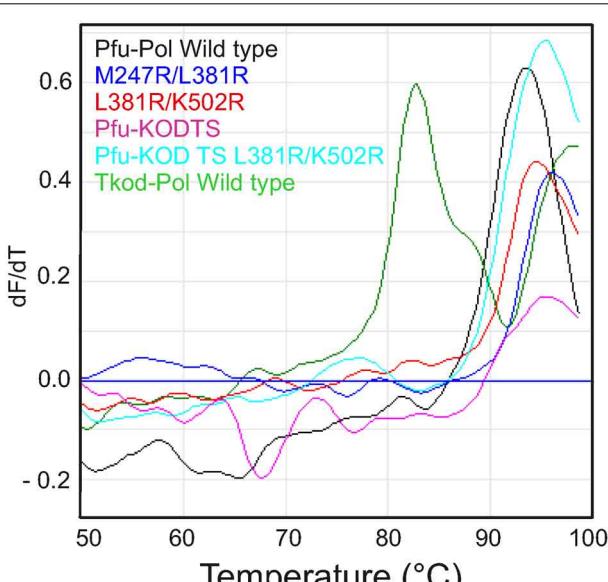


FIGURE 8 | Differential scanning fluorimetry (DSF) profiles describing the thermal unfolding of the polymerases. The first derivatives of the DSF profiles are shown with dF/dT indicating the change in fluorescence (relative units). Individual polymerases are identified by the color coding given in the figures.

the DNA phosphate backbone and disrupting conserved protein-DNA interactions (Kim et al., 2008). Deletion of this leucine did result in slightly improved primer-template extension, but it is not clear if this mutation correctly lines up the subsequent arginine in Pfu-Pol with R381 of Tgo-Pol (Figure 1) and the L381R substitution was slightly superior (supplementary data, Figure S2). Of course Pfu-PolL381R still retains an extra amino acid and this may disrupt the downstream Y-GG/A motif as seen for the wild type. Finally communication between the exonuclease and thumb domains has been implicated in the co-ordination of polymerase and proof-reading exonuclease activities (Kuroita et al., 2005; Kim et al., 2008). Trans-domain interactions between amino acids in these regions appear to control the relative motions of the two domains, influencing catalytic activities. The inter-domain contacts appear to vary slightly between Tkod-Pol and Pfu-Pol, potentially impinging on PCR properties. Such considerations may be especially significant with the thumb swap mutants.

The significance of this publication is the demonstration that it is relatively simple to improve the PCR performance of Pfu-Pol by substituting individual amino acids and even an entire domain found in the closely related Tkod-Pol. The Pfu-Pol framework seems remarkably tolerant to substantial alteration, maintaining thermostability and fidelity. Future possibilities include incorporating elements from other *Thermococcales* family-B polymerases; a reasonable number are known (Figure 1) and many have been applied in the PCR (Terpe, 2013). Further improvement may be achieved by combining gene segments using random techniques such as DNA shuffling (Stemmer, 1994) or staggered extension process (StEP) (Zhao et al., 1998) and selecting for improved PCR performance with compartmentalized self-replication (CSR) (Ghadessy and Holliger, 2007). This approach may even be extending to environmental DNA from uncharacterized archaea, widening the genetic resource (Matsukawa et al., 2009). Experiments could target the “forked-point” and thumb, as well as other elements, suggested above to play a role in PCR ability. Such an undertaking should yield important clues about features of archaeal DNA polymerases that are important for efficient and robust PCR and may, ultimately, lead to superior reagents. It may be possible to develop DNA polymerases more suited to demanding PCR applications such as amplifying DNA from single cells or old/damaged/degraded sources and in the steps needed prior to high throughput DNA sequencing.

AUTHOR CONTRIBUTIONS

All authors carried out experiments to acquire data and participated in analysing and interpreting the results. Bernard A. Connolly originally conceived the project and Ashraf M. Elshawadfy, Brian J. Keith, and Thomas Kinsman contributed to experimental design. Bernard A. Connolly wrote the first draft of the work and all authors contributed in critical revision. All authors concur with the final version and agree to take responsibility for the work and conclusions described within.

ACKNOWLEDGMENTS

Ashraf M. Elshawadfy was supported by grants from the Egyptian government and the University of Zagazig (Zagazig, Egypt). Brian J. Keith and Thomas Kinsman were UK BBSRC-supported Ph.D.

students. The research was supported by a Wellcome Trust (UK) equipment grant (grant number 064345). Zvi Kelman is thanked for kindly supplying the Tkod-Pol B overexpressing plasmid.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00224/abstract>

REFERENCES

- Aller, P., Duclos, S., Wallace, S. S., and Doublie, S. (2011). A crystallographic study of the role of sequence context in thymine glycol bypass by a replicative DNA polymerase serendipitously sheds light on the exonuclease complex. *J. Mol. Biol.* 412, 22–34. doi: 10.1016/j.jmb.2011.07.007
- Bambara, R. A., Fay, P. J., and Mallaberry, L. M. (1995). Methods of analysing processivity. *Meth. Enzymol.* 262, 270–280. doi: 10.1016/0076-6879(95)62023-0
- Bergen, K., Betz, W., Welte, W., Diederichs, K., and Marx, A. (2013). Structures of KOD and 9°N DNA polymerases complexed with primer template duplex. *Chembiochem* 14, 1058–1062. doi: 10.1002/cbic.201300175
- Bohlke, K., Pisani, F. M., Vorgias, C. E., Frey, B., Sobek, H., Rossi, M., et al. (2000). PCR performance of the B-type DNA polymerase from the thermophilic euryarchaeon *Thermococcus aggregans* improved by mutations in the Y-GG/A motif. *Nucleic Acids Res.* 28, 3910–3917. doi: 10.1093/nar/28.20.3910
- Borges, N., Matsumi, R., Imanaka, T., Atomi, H., and Santos, H. (2010). *Thermococcus kodakarensis* mutants deficient in di-myoinositol phosphate use aspartate to cope with heat stress. *J. Bacteriol.* 192, 191–197. doi: 10.1128/JB.01115-09
- Brautigam, C. A., and Steitz, T. A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* 8, 54–63. doi: 10.1016/S0959-440X(98)80010-9
- Chapin-Rodriguez, A., Park, H.-W., Mao, C., and Beese, L. S. (2000). Crystal structure of a Pol α family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9°N-7. *J. Mol. Biol.* 299, 447–462. doi: 10.1006/jmbi.2000.3728
- Cline, J., Braman, J. C., and Hogrefe, H. H. (1996). PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* 18, 3546–3551. doi: 10.1093/nar/24.18.3546
- Emptage, K., O'Neill, R., Solovyova, A., and Connolly, B. A. (2008). Interplay between DNA polymerase and proliferating cell nuclear antigen switches off base excision repair of uracil and hypoxanthine during replication in archaea. *J. Mol. Biol.* 383, 762–771. doi: 10.1016/j.jmb.2008.08.018
- Evans, S. J., Fogg, M. J., Mamone, A., Davis, M., Pearl, L. H., and Connolly, B. A. (2000). Improving dideoxynucleotide-triphosphate utilization by the hyperthermophilic DNA polymerase from *Pyrococcus furiosus*. *Nucleic Acids Res.* 28, 1059–1066. doi: 10.1093/nar/28.5.1059
- Fiala, G., and Stetter, K. O. (1986). *Pyrococcus furiosus*, sp. nov., represents a novel genus of marine heterotrophic archaeabacteria growing optimally at 100°C. *Arch. Microbiol.* 145, 56–61. doi: 10.1007/BF00413027
- Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J., and Connolly, B. A. (2008). Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.* 381, 529–539. doi: 10.1016/j.jmb.2008.06.004
- Ghadessy, F. J., and Holliger, P. (2007). Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes. *Methods Mol. Biol.* 352, 237–248.
- Gougel, J., Ralec, C., Henneke, G., and Delarue, M. (2012). Molecular recognition of canonical and deaminated bases by *P. abyssi* family B DNA polymerase. *J. Mol. Biol.* 423, 315–336. doi: 10.1016/j.jmb.2012.07.025
- Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., et al. (2001). Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J. Mol. Biol.* 306, 469–477. doi: 10.1006/jmbi.2000.4403
- Hogg, M., Ostermal, P., Bylund, G. O., Ganail, R. A., Lundström, E.-B., Sauer-Eriksson, A. E., et al. (2014). Structural basis for processive DNA synthesis by yeast DNA polymerase ε. *Nat. Struct. Mol. Biol.* 21, 49–55. doi: 10.1038/nsmb.2712
- Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., et al. (1999). Crystal structure of a thermostable type B DNA polymerase from

- Thermococcus gorgonarius*. Proc. Natl. Acad. Sci. U.S.A. 96, 3600–3605. doi: 10.1073/pnas.96.7.3600
- Ishino, S., Kawamura, A., and Ishino, Y. (2012). Application of PCNA to processive PCR by reducing the stability of its ring structure. J. Jpn. Soc. Extremophiles 11, 19–25.
- Ishino, Y., and Ishino, S. (2011). Rapid progress of DNA replication studies in Archaea, the third domain of life. China Life Sci. 55, 386–403. doi: 10.1007/s11427-012-4324-9
- Jozwiakowski, S. J., and Connolly, B. A. (2009). Plasmid-based *lacZα* assay for DNA polymerase fidelity: application to archaeal family-B DNA polymerase. Nucleic Acids Res. 37, e102. doi: 10.1093/nar/gkp494
- Keith, B. J., Jozwiakowski, S. K., and Connolly, B. A. (2013). A plasmid-based *lacZα* gene assay for DNA polymerase fidelity measurement. Anal. Biochem. 433, 153–161. doi: 10.1016/j.ab.2012.10.019
- Killelea, T., and Connolly, B. A. (2011). Role of disulfide bridges in archaeal family-B DNA polymerases. Chembiochem 12, 1330–1336. doi: 10.1002/cbic.201100145
- Killelea, T., Ghosh, S., Tan, S. S., Heslop, P., Firbank, S., Kool, E. T., et al. (2010). Probing the interaction of archaeal DNA polymerases with deaminated bases using X-ray crystallography and non-hydrogen bonding isosteric base analogues. Biochemistry 49, 5772–5781. doi: 10.1021/bi100421r
- Kim, S. W., Kim, D.-U., Kim, J. K., Kang, L.-W., and Cho, H.-S. (2008). Crystal structure of *Pfu*, the high fidelity DNA polymerase from *Pyrococcus furiosus*. Int. J. Biol. Macromol. 42, 356–361. doi: 10.1016/j.ijbiomac.2008.01.010
- Kim, Y. J., Lee, H. S., Bae, S. S., Jeon, J. H., Lim, J. K., Cho, Y., et al. (2007). Cloning, purification, and characterization of a new DNA polymerase from a hyperthermophilic archaeon, *Thermococcus* sp. NA1. J. Microbiol. Biotechnol. 17, 1090–1097.
- Kitabayashi, M., Nishiya, Y., Esaka, M., Itakura, M., and Imanaka, T. (2002). Gene cloning and polymerase chain reaction with proliferating cell nuclear antigen from *Thermococcus kodakarensis* KOD1. Biosci. Biotechnol. Biochem. 66, 2194–2200. doi: 10.1271/bbb.66.2194
- Kuroita, T., Matsumura, H., Yokota, N., Kitabayashi, M., Hashimoto, H., Inoue, T., et al. (2005). Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases. J. Mol. Biol. 351, 291–298. doi: 10.1016/j.jmb.2005.06.015
- Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., and Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. Gene 108, 1–6. doi: 10.1016/0378-1119(91)90480-Y
- Matsukawa, H., Yamagami, T., Kawarabayashi, Y., Miyashita, Y., Takahashi, M., and Ishino, Y. (2009). A useful strategy to construct DNA polymerases with different properties by using genetic resources from environmental DNA. Genes Genet. Syst. 84, 3–13. doi: 10.1266/ggs.84.3
- McPherson, M. J., and Möller, S. G. (2006). PCR. Boca Raton, FL: Taylor and Francis.
- Mullis, K. B., Ferre, F., and Gibbs, R. A. (1994). *The Polymerase Chain Reaction (PCR)*. Basel: Birkhäuser Verlag AG. doi: 10.1007/978-1-4612-0257-8
- Nishioka, M., Mizuguchi, H., Fujiwara, S., Komatsubara, S., Kitabayashi, M., Uemura, H., et al. (2001). Long and accurate PCR with a mixture of KOD DNA polymerase and its exonuclease deficient mutant enzyme. J. Biotech. 88, 144–149. doi: 10.1016/S0168-1656(01)00275-9
- Pavlov, A. R., Belova, G. I., Kozyavkin, S. A., and Slesarev, A. I. (2002). Helix-turn-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. Proc. Natl. Acad. Sci. U.S.A. 99, 13510–13515. doi: 10.1073/pnas.202127199
- Reha-Krantz, L. J. (2010). DNA polymerase proofreading: Multiple roles maintain genome stability. Biochimica Biophysica Acta 1804, 1049–1063. doi: 10.1016/j.bbapap.2009.06.012
- Richardson, T. T., Wu, X., Keith, B. J., Heslop, P., Jones, A. C., and Connolly, B. A. (2013). Unwinding of primer-templates by archaeal family-B DNA polymerases in response to template-strand uracil. Nucleic Acids Res. 41, 2466–2478. doi: 10.1093/nar/gks1364
- Russell, H. J., Richardson, T. T., Emptage, K., and Connolly, B. A. (2009). The 3'-5' proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template strand deaminated bases. Nucleic Acids Res. 37, 7603–7611. doi: 10.1093/nar/gkp800
- Saunders, N. A., and Lee, M. (2013). *Real Time PCR; Advanced Technologies and Applications*. Linton: Caister Academic Press.
- Shamoo, Y., and Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell 99, 155–166. doi: 10.1016/S0092-8674(00)81647-5
- Shuttleworth, G., Fogg, M. J., Kurpiewski, M. R., Jen-Jacobson, L., and Connolly, B. A. (2004). Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea. J. Mol. Biol. 337, 621–634. doi: 10.1016/j.jmb.2004.01.021
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539. doi: 10.1038/msb.2011.75
- Stemmer, C. P. (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. Nature 370, 389–391. doi: 10.1038/370389a0
- Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009). Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ. Nat. Struct. Mol. Biol. 16, 979–986. doi: 10.1038/nsmb.1663
- Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., et al. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. Appl. Environ. Microbiol. 63, 4504–4510.
- Terpe, K. (2013). Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 97, 10253–10254. doi: 10.1007/s00253-013-5290-2
- Von Hippel, P. H., Fairfield, F. R., and Dolejsi, M. K. (1994). On the processivity of polymerases. Ann. N.Y. Acad. Sci. 726, 118–131. doi: 10.1111/j.1749-6632.1994.tb52803.x
- Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. Nucleic Acids Res. 32, 1197–1207. doi: 10.1093/nar/gkh271
- Warrens, A. N., Jones, M. D., and Lechner, R. I. (1997). Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. Gene 186, 29–35. doi: 10.1016/S0378-1119(96)00674-9
- Weissensteiner, T., Griffin, H. G., and Griffin, A. M. (2004). *PCR Technology: Current Innovations*. Boca Raton, FL: CRC Press.
- Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nat. Biotechnol. 16, 258–261. doi: 10.1038/nbt0398-258
- Conflict of Interest Statement:** Brian J. Keith was partially funded by Bioline, a company with commercial interests in PCR enzymes. Apart from this the authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 04 April 2014; **accepted:** 28 April 2014; **published online:** 27 May 2014.
- Citation:** Elshawadfy AM, Keith BJ, Ee Ooi H, Kinsman T, Heslop P and Connolly BA (2014) DNA polymerase hybrids derived from the family-B enzymes of *Pyrococcus furiosus* and *Thermococcus kodakarensis*: improving performance in the polymerase chain reaction. Front. Microbiol. 5:224. doi: 10.3389/fmicb.2014.00224
- This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.
- Copyright © 2014 Elshawadfy, Keith, Ee Ooi, Kinsman, Heslop and Connolly. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering

Takeshi Yamagami¹, Sonoko Ishino¹, Yutaka Kawarabayashi^{1,2} and Yoshizumi Ishino^{1*}

¹ Protein Chemistry and Engineering, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka, Japan

² Health Research Institute, National Institute of Advanced Industrial Science and Technology, Amagasaki, Japan

Edited by:

Zvi Kelman, University of Maryland, USA

Reviewed by:

Paul Beare, National Institutes of Health, USA

Zhuo Li, State Oceanic Administration, China

***Correspondence:**

Yoshizumi Ishino, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka, Fukuoka 812-8581, Japan
e-mail: ishino@agr.kyushu-u.ac.jp

DNA polymerases are widely used for DNA manipulation *in vitro*, including DNA cloning, sequencing, DNA labeling, mutagenesis, and other experiments. Thermostable DNA polymerases are especially useful and became quite valuable after the development of PCR technology. A DNA polymerase from *Thermus aquaticus* (Taq polymerase) is the most famous DNA polymerase as a PCR enzyme, and has been widely used all over the world. In this study, the gene fragments of the family A DNA polymerases were amplified by PCR from the DNAs from microorganisms within environmental soil samples, using a primer set for the two conserved regions. The corresponding region of the *pol* gene for Taq polymerase was substituted with the amplified gene fragments, and various chimeric DNA polymerases were prepared. Based on the properties of these chimeric enzymes and their sequences, two residues, E742 and A743, in Taq polymerase were found to be critical for its elongation ability. Taq polymerases with mutations at 742 and 743 actually showed higher DNA affinity and faster primer extension ability. These factors also affected the PCR performance of the DNA polymerase, and improved PCR results were observed with the mutant Taq polymerase.

Keywords: thermostability, gene amplification, *in vitro* gene manipulation, *Thermus aquaticus*, PCR

INTRODUCTION

In addition to their fundamental roles in maintaining genome integrity during replication and repair, DNA polymerases are widely used for genetic engineering techniques, including DNA cloning, dideoxy-sequencing, DNA labeling, mutagenesis, and other *in vitro* DNA manipulations. Among them, thermostable DNA polymerases are particularly useful for PCR and cycle-sequencing (Perler et al., 1996; Ishino and Ishino, 2013; Terpe, 2013).

The fundamental ability to synthesize a deoxyribonucleotide chain is conserved in relation to the structural conservation of the DNA polymerases. However, the more specific properties for this catalysis, including processivity, synthesis accuracy, and substrate nucleotide selectivity, differ among the enzymes. These factors should be considered when evaluating a DNA polymerase as an enzyme for genetic engineering (Ishino and Ishino, 2014). An enzyme possessing faster extension with better accuracy and higher efficiency is more preferable. In addition to these catalytic properties, thermostability is necessary for practical PCR. DNA polymerases are now classified into seven families, based on the amino acid sequences (Braithwaite and Ito, 1993; Ishino and Cann, 1998; Cann and Ishino, 1999; Ohmori et al., 2001; Lipps et al., 2003). The enzymes within the same family have basically similar properties. Commercial genetic engineering reagents have originated only from families A and B to date. The family A enzymes are used for dideoxy-sequencing, and the family A and B enzymes are used for PCR. None of the DNA polymerases

from the other families are suitable for general use in genetic engineering experiments. The 3'-5' exonuclease activity, which contributes to the proofreading of DNA strand synthesis, is generally associated with the family B enzymes, but not with the family A enzymes, although some family A enzymes have a weak 3'-5' exonuclease activity (Joyce and Steitz, 1994; Villbrandt et al., 2000). Based on these differences, family A is advantageous for the efficient amplification of a long DNA region, and family B is generally more suitable for the precise amplification of a shorter region by PCR (Eckert and Kunkel, 1991). Researchers in this field have been making continuous efforts toward longer extension and better accuracy in PCR, and have succeeded in developing practical and reliable PCR methods. One notable example is the development of LA (long and accurate)-PCR, which is performed with a mixture of two DNA polymerases, one each from family A and family B (Barns, 1994). Further improvements of PCR have included the identification of a processive enzyme (Takagi et al., 1997) and the modifications within family B DNA polymerases that confer higher accuracy (Wang et al., 2004; Ishino et al., 2012).

Protein engineering techniques, using site-specific or random mutagenesis, are powerful ways to create mutant enzymes from the known DNA polymerases. Several useful enzymes were successfully produced by these procedures. A cold-sensitive mutant of DNA polymerase from *Thermus aquaticus* (Taq polymerase) was developed with markedly reduced activity at 37°C, as compared with the wild type (WT) enzyme (Kermekchiev et al.,

2003). This mutant may be applicable to hot start PCR. Another example is a mutant Taq polymerase with enhanced resistance to various inhibitors of PCR reactions, including whole blood, plasma, hemoglobin, lactoferrin, serum IgG, soil extracts, and humic acid (Kermekchiev et al., 2009). The molecular breeding of *Thermus* DNA polymerases by a direct evolution technique (Brakmann, 2005; Henry and Romesberg, 2005; Holmberg et al., 2005; Ong et al., 2006), compartmentalized self-replication (CSR) (Ghadessy et al., 2001), also generated a PCR enzyme with striking resistance to a wide range of inhibitors (Baar et al., 2011). Furthermore, enzymes with a broad substrate specificity spectrum were also obtained by the CSR technique (Ghadessy et al., 2004; d'Abbadie et al., 2007), and are thus useful for the amplification of ancient DNA containing numerous lesions. Mutational studies in the O-helix of Taq polymerase produced enzymes with reduced fidelity (Suzuki et al., 1997, 2000; Tosaka et al., 2001), which may be useful for error-prone PCR. These studies have contributed to the elucidation of the detailed structure-function relationships of DNA polymerases, as well as to the creation of novel enzymes with different substrate specificities, stabilities, and activities from those of their naturally evolved counterparts.

In addition to the engineering of characterized enzymes to convert PCR performance, the screening for a suitable DNA polymerase activity from known organisms is the most conventional way to discover useful enzymes. However, the culturable organisms are limited, and large-scale cultivation is needed to purify an enzyme to homogeneity for precise characterization. In this study, we analyzed the DNAs from microorganisms within various soil samples obtained from a hot spring area, and compared the sequences of a region within the *pol* genes included in the environmental DNAs. We then predicted the amino acid residues that are critical for the primer extension reaction of Taq polymerase, by constructing numerous chimeric Taq polymerases including the *pol* gene fragments from the various environmental DNAs. A mutant Taq polymerase with the E742 and A743 substitutions possessed more efficient DNA strand synthesis ability and better PCR performance. This polymerase will contribute to the development of high-speed PCR with the standard PCR conditions optimized for Taq polymerase.

MATERIALS AND METHODS

ENZYMES AND SUBSTRATES

Enzymes for *in vitro* DNA manipulation and oligonucleotides were purchased from New England Biolabs (Ipswich, MA, USA) and Sigma Aldrich (St. Louis, MO, USA), respectively. The [³H]TTP was purchased from Amersham (Buckinghamshire, UK) and the [³²P]ATP was purchased from NEN Life Science Products (Boston, MA, USA).

DNA EXTRACTION

Extraction of DNA from the environmental specimens was performed with an UltraClean Soil DNA Isolation Kit (MO BIO, San Diego, CA, USA), according to the manufacturer's instructions. The extracted DNAs were assessed by agarose gel electrophoresis and were quantified by spectrophotometrical measurement.

CONSTRUCTION OF THE EXPRESSION PLASMID FOR CHIMERIC Taq POLYMERASES

The expression plasmid for Taq polymerase, pTV-Taq, which contains the entire region of the structural gene encoding Taq polymerase in the pTV118N vector (Takara Bio, Shiga, Japan), was constructed exactly as described (Ishino et al., 1994). The gene in the pTV118N vector was expressed under control of the *lac* promoter and SD sequence. The pTV-Taq plasmid was subjected to site-specific mutagenesis, using QuikChange™ kit (Agilent Technologies, Santa Clara, CA, USA) to introduce BpI (GCTNAGC) and BgIII (AGATCT) restriction sites into the positions corresponding to the 5'- and 3'-termini, respectively, of the substitution region within the Taq *pol* gene. The insertion of a BgIII site leads to the mutations of amino acids, Leu787Ile and Val788Leu. The resultant plasmid was named pTV-Taq', and it was used for the expression of chimeric Taq polymerases, produced by the direct substitution of the BpI-BgIII fragment with the *pol* gene fragments from the environmental DNAs, as described in detail below.

AMPLIFICATION OF THE *pol* GENE FRAGMENTS FROM METAGENOMIC DNA

A region of the *pol* genes was amplified directly from the environmental DNA by PCR, using a primer set with the sequences 5'-dC GCAGGCTAAGCAGCTCCGAYCCHAACYTSCARAAYATHCC-3' and 5'- dGAGYA**AGATCT**CRTCGTGNACYTG-3', which correspond to the degenerate codons for DPNLQNIP (forward) and QVHDEIL (reverse), respectively (Y indicates C and T; H indicates A, C, and T; S indicates C and G; R indicates A and G; N indicates A, G, C, and T). The above two regions, which are conserved in the family A DNA polymerases, were successfully used to make a mixed primer set for PCR (Uemori et al., 1993). The nucleotide sequences corresponding to the conserved regions are shown in boldface, and the restriction endonuclease recognition sequences are underlined in the above primers. PCR was performed in a 50 µl reaction, containing 10 ng DNA, 25 pmol of each primer, 0.2 mM dNTP, and 1 unit of PfuUltra DNA polymerase (Stratagene). After an incubation of the mixture without the enzyme for 3 min at 95°C, thirty-cycles of PCR, with a temperature profile of 30 s at 95°C, 30 s at 55°C, and 1 min at 72°C, were performed. The reaction mixtures were electrophoresed in a 1% agarose gel, and the amplified fragments were visualized by ethidium bromide staining.

ANALYSIS OF THE AMPLIFIED GENE FRAGMENTS

DNA fragments with a 600 bp size, amplified from the environmental DNA by PCR, were excised from the agarose gel, digested with the BpI and BgIII restriction endonucleases, and ligated into the pTV-Taq' plasmid predigested with the BpI and BgIII restriction endonucleases. The ligation mixtures were introduced into *E. coli* JM109 cells (TaKaRa Bio.), and 20 clones were picked independently from the transformants for each amplification. Plasmid DNAs were extracted from these clones, and the nucleotide sequences of the DNA inserts were determined by CEQ2000XL DNA analysis system (Beckman Coulter, USA).

CONSTRUCTION OF MUTANT Taq POLYMERASES

The pTV-Taq plasmid was subjected to site-specific mutagenesis to introduce mutations into positions 742 and 743 of Taq polymerase. The sequences of the 14 primers used for mutagenesis are shown in **Table 1**. The PCR reaction mixture contained 20 ng template DNA, 1× PCR buffer for KOD-Plus-Neo, 1.5 mM Mg₂SO₄, 0.2 mM of each dNTP, 0.3 μM of each primer, and 0.5 unit KOD-Plus-Neo DNA polymerase (TOYOBO, Osaka, Japan) in a final volume of 20 μl. The mixture was heated at 95°C for 30 s and then subjected to thermal cycling (14 cycles of 95°C for 30 s, 55°C for 1 min, and 68°C for 8 min). The PCR product was treated with DpnI at 37°C for 1 h, and introduced into *E. coli* JM109 cells. For each mutation, the polymerase gene was fully sequenced to ensure that the mutation of interest was present and that no additional mutation was introduced by the PCR.

PURIFICATION OF WILD TYPE AND MUTANT DNA POLYMERASES

E. coli JM109 cells carrying the expression plasmid were grown at 37°C, in 1 L of LB medium containing 100 μg/ml ampicillin.

Table 1 | Oligonucleotides used to introduce mutations into positions 742 and 743 of Taq polymerase.

Name of primer	Sequence (5' to 3')
TaqRR-F	CGGGTGAAGAGCGTGC <u>GGCCGCGCCGAGCGCATGGCC</u>
TaqRR-R	GGCATGCGCT <u>CGGCCGCGCGCACGCTTCA</u> CCCG
TaqRA-F	CGGGTGAAGAGCGTGC <u>GGCGCGCCGAGCGCATGGCC</u>
TaqRA-R	GGCATGCGCT <u>CGGCCGCGCGCACGCTTCA</u> CCCG
TaqAA-F	CGGGTGAAGAGCGTGC <u>GGCGGGCCGAGCGCATGGCC</u>
TaqAA-R	GGCATGCGCT <u>CGGCCGCGCGCACGCTTCA</u> CCCG
TaqER-F	CGGGTGAAGAGCGTGC <u>GCGAGCGCGCACGCTTCA</u> CCCG
TaqER-R	GGCATGCGCT <u>CGGCCGCGCGCACGCTTCA</u> CCCG
TaqAR-F	CGGGTGAAGAGCGTGC <u>CGCGCGCCGAGCGCATGGCC</u>
TaqAR-R	GGCATGCGCT <u>CGGCCGCGCGCACGCTTCA</u> CCCG
TaqRK-F	CGGGTGAAGAGCGTGC <u>CGCAAAGCGAGCGCATGGCC</u>
TaqRK-R	GGCATGCGCT <u>CGGCGCGCACGCTTCA</u> CCCG
TaqKR-F	CGGGTGAAGAGCGTGC <u>CGAACACGCCGAGCGCATGGCC</u>
TaqKR-R	GGCATGCGCT <u>CGGCGCGCGCACGCTTCA</u> CCCG
TaqKK-F	CGGGTGAAGAGCGTGC <u>CGAAAAAGCGAGCGCATGGCC</u>
TaqKK-R	GGCATGCGCT <u>CGGCGCGCACGCTTCA</u> CCCG
TaqQY-F	CGGGTGAAGAGCGTGC <u>GCAGTATGCCGAGCGCATGGCC</u>
TaqQY-R	GGCATGCGCT <u>CGGCATACTGGCGCACGCTTCA</u> CCCG
TaqAH-F	CGGGTGAAGAGCGTGC <u>CGCGCATGCCGAGCGCATGGCC</u>
TaqAH-R	GGCATGCGCT <u>CGGCATGCCGCGCACGCTTCA</u> CCCG
TaqEH-F	CGGGTGAAGAGCGTGC <u>CGCGAGCATGCCGAGCGCATGGCC</u>
TaqEH-R	GGCATGCGCT <u>CGGCATGCCGCGCACGCTTCA</u> CCCG
TaqHA-F	CGGGTGAAGAGCGTGC <u>CGCGCATGCCGAGCGCATGGCC</u>
TaqHA-R	GGCATGCGCT <u>CGGCCATGGCGCACGCTTCA</u> CCCG
TaqHH-F	CGGGTGAAGAGCGTGC <u>CGCCATCATGCCGAGCGCATGGCC</u>
TaqHH-R	GGCATGCGCT <u>CGGCATGCCGAGCGCATGGCC</u>
TaqHK-F	CGGGTGAAGAGCGTGC <u>GCCTAAAGCGAGCGCATGGCC</u>
TaqHK-R	GGCATGCGCT <u>CGGCTTATGCCGAGCGCATGGCC</u>

The sites for the mutations are underlined.

The cells were cultured to an A₆₀₀ of 0.2–0.3, and then the expression of the *pol* gene was induced by further cultivation for 3 h in the presence of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were harvested and disrupted by sonication in the lysis solution (50 mM Tris-HCl, pH 8.0, 1 mM DTT, 1 mM EDTA, 1 mM PMSF). For the preparation of the WT and mutant Taq polymerases, the soluble cell extract, obtained by centrifugation at 12,000 × g for 20 min, was heated at 75°C for 30 min. The heat-stable fraction was obtained by centrifugation, and was treated with 0.15% (w/v) polyethyleneimine in the presence of 1 M NaCl, to remove the nucleic acids. The soluble proteins were precipitated by 80%-saturated ammonium sulfate. The precipitate was resuspended in the buffer [50 mM Tris-HCl, pH 8.0, 0.5 M (NH₄)₂SO₄], and was subjected to chromatography on a hydrophobic column (HiTrap Phenyl HP 5 ml, GE Healthcare). The column was washed with 50 mM Tris-HCl, pH 8.0, and the Taq polymerase was eluted with deionized water. An equal volume of 100 mM Tris-HCl, pH 8.0, was added to this fraction, and it was then subjected to affinity chromatography (HiTrap Heparin HP 5 ml, GE Healthcare) with a gradient of 0–2 M NaCl. The proteins that eluted at 0.8 M NaCl were stored at –25°C as the final sample, in 20 mM Tris-HCl, pH 8.0, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% Tween 20, and 50% (w/v) glycerol.

MEASUREMENT OF NUCLEOTIDE INCORPORATION ACTIVITY

The DNA polymerizing activity was assayed by measuring the incorporation of [³H]TTP into the acid insoluble materials, basically as described previously (Uemori et al., 1995). To a 50 μl solution, containing 20 mM Tris-HCl, pH 8.8, 5 mM MgCl₂, 14 mM 2-mercaptoethanol, 0.2 mM each dATP, dGTP, dCTP, and dTTP, 400 nM [³H]dTTP, and 20 μg of activated salmon sperm DNA, a constant amount of the enzyme fraction was added, and the reaction was incubated at 74°C for 2.5, 5, and 10 min. After the reaction, a 10 μl portion of each reaction mixture was spotted onto DE81 filters (GE Healthcare Japan, Tokyo, Japan). The filters were washed three times with a 5% Na₂HPO₄ solution, and the radioactivity incorporated into the DNA strands was counted by a scintillation counter. One unit of activity is defined as the amount of enzyme catalyzing the incorporation of 10 nmol of dNTP into DNA per 30 min at 74°C, and the specific activity was calculated as the units per one mg of protein (units/mg) for each DNA polymerase.

PRIMER EXTENSION ACTIVITY

The primer extension ability was investigated by using M13 single-stranded DNA (ssDNA) annealed with a ³²P-labeled DNA, as described previously (Cann et al., 1999). M13 ssDNA (0.05 pmol), annealed with a 55 nucleotide long primer (5'-dT_C GTAAATCATGGTCATAGCTGTTCTGTGAAATTGTTATCC GCTCACAAATTC-3'), was mixed with each DNA polymerase (0.05 unit) and dNTP (0.2 mM) in a 20 μl solution, containing 20 mM Tris-HCl, pH 8.8, 5 mM MgCl₂, and 14 mM 2-mercaptoethanol, and incubated 74°C for 5 min. The reaction mixtures were analyzed by alkaline agarose gel electrophoresis, and the sizes of the products were visualized by autoradiography.

ELECTROPHORETIC MOBILITY-SHIFT ASSAY

The electrophoretic mobility-shift assay (EMSA) was performed as described previously (Komori and Ishino, 2000), to measure the DNA binding ability of the DNA polymerases. The ^{32}P -labeled 27mer oligonucleotide (5'-dAGCTATGACCATGATTACGAATTGCTT-3') was annealed with the 49mer oligonucleotide (5'-dA GCTACCAGCCTGCACGAATTAAAGCAATTCTGTAATCATGGTCATAGCT-3') and was used as a DNA substrate. The radiolabeled DNA (3 nM) was mixed with DNA polymerase proteins (0.6–400 nM) in a 20 μl solution, containing 20 mM Tris-HCl, pH 8.8, 10 mM NaCl, 5 mM MgCl₂, 14 mM 2-mercaptoethanol, 0.1 mg/ml BSA, and 5% (w/v) glycerol, and incubated at 40°C for 5 min. The DNA-enzyme mixtures were fractionated by 1% agarose gel electrophoresis. The autoradiograms were scanned and the band intensities were quantified using an image analyzer (FLA5000; Fuji Film, Tokyo, Japan). The fraction of bound DNA in each lane was calculated to be: fraction bound DNA = 1 – fraction unbound DNA. The quantitated data for binding (association) was plotted vs. enzyme concentrations. The apparent K_d was determined to be the protein concentration at which the degree of binding equals 0.5.

RESULTS

AMPLIFICATION OF FAMILY A DNA POLYMERASE GENE

We collected soil samples from various hot spring areas in Japan, and isolated the DNAs within these samples. Using these DNAs as templates, a region of the *pol* genes encoding family A DNA polymerase-like sequences was amplified. We previously reported the amplification of the *pol* gene encoding a family A DNA polymerase by a set of mixed primers based on the two conserved sequence motifs (Uemori et al., 1993). The mixed primer set worked for the specific amplification of the *pol* gene fragments from the hot spring samples in this study. As shown in **Figure 1** (a part of the experiments is shown), a 600 bp DNA fragment was amplified as a single band from several samples. We tested 384 different samples, and detected the amplified DNA in 37 samples, obtained at Onikobe, Hachimantai, Nasu, Kirishima, and Beppu, as shown in **Table 2**. These locations had various environmental conditions, including pH values of 1~7 and temperatures

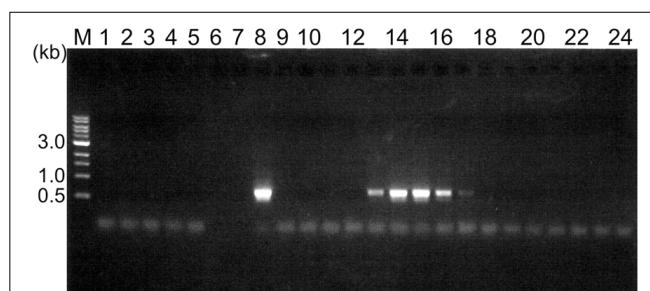


FIGURE 1 | Amplification of a region within the family A DNA polymerase gene from the environmental DNA. PCR reaction mixtures were fractionated by 1% agarose gel electrophoresis, and the DNA bands were visualized by ethidium bromide staining. The lane numbers represent the serial numbers of the samples obtained from the hot spring areas in the Tohoku and Kyushu districts.

of mostly 70–100°C, and thus were expected to be inhabited by microorganisms with highly diverse genetic resources. The efficiency of DNA isolation was not addressed in this experiment, and thus it is possible that a sufficient amount of DNA was not present in some samples. Therefore, the result that only about 10% of the samples provided target gene amplification is not directly related to the presence or absence of microorganisms in the samples. The amplified gene fragments were excised from the gel, and were cloned into a plasmid vector. Twenty colonies were isolated independently from each cloning of the amplified DNA, and a total of 740 (20 \times 37 amplifications) plasmids were isolated. These cloned DNA fragments were subjected to

Table 2 | Summary of metagenomic analyses.

No.	Sampling place (area)	Temperatuare (°C)	pH	Number of clones	Number of different clone
1	Koyasukyo (Tohoku)	98	5	20	5
2	Koyasukyo (Tohoku)	70	7	20	12
3	Koyasukyo (Tohoku)	82	6.5	20	6
4	Koyasukyo (Tohoku)	73	—	20	16
5	Koyasukyo (Tohoku)	74	6.7	20	2
6	Koyasukyo (Tohoku)	35	—	20	10
7	Koyasukyo (Tohoku)	82	7	20	13
8	Koyasukyo (Tohoku)	50	8	20	12
9	Koyasukyo (Tohoku)	57	7	20	11
10	Koyasukyo (Tohoku)	96	4	20	4
11	Onikobe (Tohoku)	94	7	20	7
12	Onikobe (Tohoku)	95	7	20	2
13	Onikobe (Tohoku)	98	8	20	2
14	Onikobe (Tohoku)	89	7	20	9
15	Onikobe (Tohoku)	96	7	20	6
16	Onikobe (Tohoku)	98	8	20	5
17	Onikobe (Tohoku)	56	7	20	9
18	Onikobe (Tohoku)	93	1	20	6
19	Onikobe (Tohoku)	50	2	20	5
20	Onikobe (Tohoku)	85	7	20	4
21	Onikobe (Tohoku)	35	1	20	8
22	Beppu (Kyushu)	96	7	20	2
23	Beppu (Kyushu)	65	6	20	7
24	Beppu (Kyushu)	44	6	20	6
25	Beppu (Kyushu)	77	6.5	20	3
26	Beppu (Kyushu)	97	5	20	6
27	Beppu (Kyushu)	70	6	20	8
28	Beppu (Kyushu)	78	7	20	5
29	Beppu (Kyushu)	—	—	20	10
30	Kirishima (Kyushu)	75	3	20	5
31	Kirishima (Kyushu)	75	5	20	9
32	Kirishima (Kyushu)	—	5	20	6
33	Kirishima (Kyushu)	—	—	20	12
34	Ibusuki (Kyushu)	94	6	20	3
35	Nasu (Kanto)	65	6.5	20	5
36	Nasu (Kanto)	55	7	20	1
37	Nasu (Kanto)	55	6.5	20	8

“—” indicates “not analyzed.”

sequencing, and the different sequences were counted (**Table 2**). In total, we obtained 250 different sequences, which were not present in the public databases. This result suggested that there are still many uncharacterized DNA polymerases in the soil samples, as expected.

PREPARATION OF THE CHIMERIC Taq POLYMERASES

The amplified gene fragments encode the region in the family A DNA polymerase that is important for the nucleotide connecting reaction. To investigate the structure and function relationships of the family A DNA polymerases, we substituted the corresponding region of Taq polymerase gene with the amplified gene fragments *in vitro*. To construct the expression plasmids for the chimeric Taq polymerases systematically, restriction sites were created at appropriate sites in the Taq *pol* structural gene (**Figure 2**). For this purpose, the *BpI* and *BgII* recognition sequences were suitable, although the substitutions of two amino acids, Leu797Val and Ile798Leu, could not be avoided by the introduction of *BgIII* at the reverse priming site. The mutant Taq polymerase

(L797V/I798L) was purified, and we confirmed that its fundamental properties were not affected. Therefore, the mutant Taq, designated as Taq' polymerase, was considered to be equivalent to the WT Taq polymerase. PCR primers containing the recognition sequences for *BpI* (forward primer) and *BgII* (reverse primer) at each 5'-terminus were synthesized and each cloned DNA was re-amplified, and thus 250 chimeric genes were constructed in the Taq polymerase expression plasmid. The total cell extracts of *E. coli* producing recombinant chimeric Taq polymerases were treated at 75°C for 30 min, and the supernatants were assayed to measure the nucleotide incorporation activity. About half of the chimeric enzymes were inactivated by the heat treatment. The thermostable chimeric Taq polymerases were further purified to apparent homogeneity by the procedure described in the Materials and Methods, and the specific activity (units/mg protein) was measured by the standard DNA polymerase assay, using activated DNA. Furthermore, the primer extension ability was evaluated for each enzyme, using a constant amount (unit). These results are summarized in **Table 3**, which shows only the chimeric Taq polymerases possessing extension abilities better than 5 kb per 5 min-reaction. As shown in the **Table 3**, 13 enzymes were superior to WT Taq polymerase. However, the thermostabilities of these high-speed DNA polymerases were not sufficient for PCR applications (data not shown).

SEQUENCE COMPARISON OF CHIMERIC Taq POLYMERASES AND CONSTRUCTION OF THE MUTANT Taq POLYMERASES

The amino acid sequences of the chimeric Taq polymerases with extension rates greater than 1 kb/min (in the condition of 0.0025 unit/μl) are aligned in **Figure 3**. In this alignment, we focused on the region from amino acids 730 to 745 in the Taq polymerase. One distinct feature is that 3 genes and 4 genes have insertions of 9 amino acids and 3 amino acids, respectively, as compared with WT Taq polymerase. The other characteristic feature is that continuous stretches of basic amino acids were

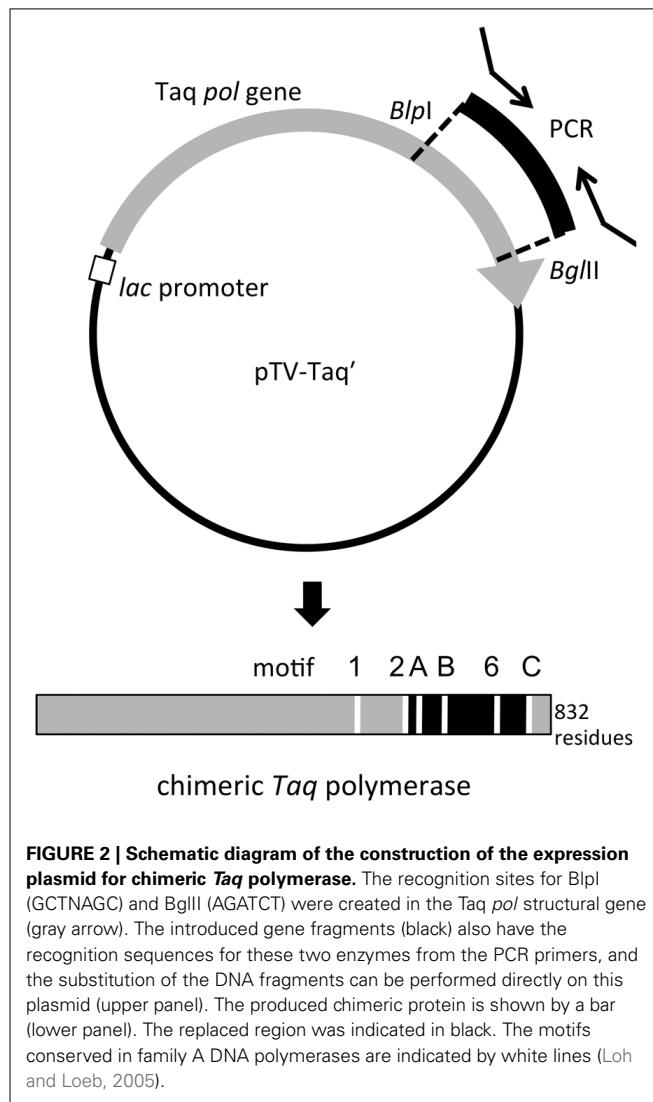


FIGURE 2 | Schematic diagram of the construction of the expression plasmid for chimeric Taq polymerase. The recognition sites for *BpI* (GCTNAGC) and *BgII* (AGATCT) were created in the Taq *pol* structural gene (gray arrow). The introduced gene fragments (black) also have the recognition sequences for these two enzymes from the PCR primers, and the substitution of the DNA fragments can be performed directly on this plasmid (upper panel). The produced chimeric protein is shown by a bar (lower panel). The replaced region was indicated in black. The motifs conserved in family A DNA polymerases are indicated by white lines (Loh and Loeb, 2005).

Table 3 | Properties of the chimeric Taq polymerase.

Name	U/mg ($\times 10^5$)	Extention kb/5 min	Kd for DNA (nM)	Sampling pace	
				Temperature (°C)	pH
TaqWT	5.1	0.7	400	–	–
8-1	1.2	5.8	10	50	8
7-1	1.2	5.8	10	82	7
29-1	0.94	6.5	10	–	–
8-2	3.3	6.0	20	50	8
7-2	2.5	5.9	5	82	7
1-1	1.1	5.8	6	98	5
30-1	1.9	6.5	8	75	3
1-2	2.1	6.4	5	98	5
10-1	1.8	6.0	4	96	4
12-1	0.73	6.4	7	95	7
4-1	0.58	6.8	4	73	–
4-2	0.51	7.5	4	73	–
7-3	0.48	7.5	4	82	7

– indicate "not analyzed."

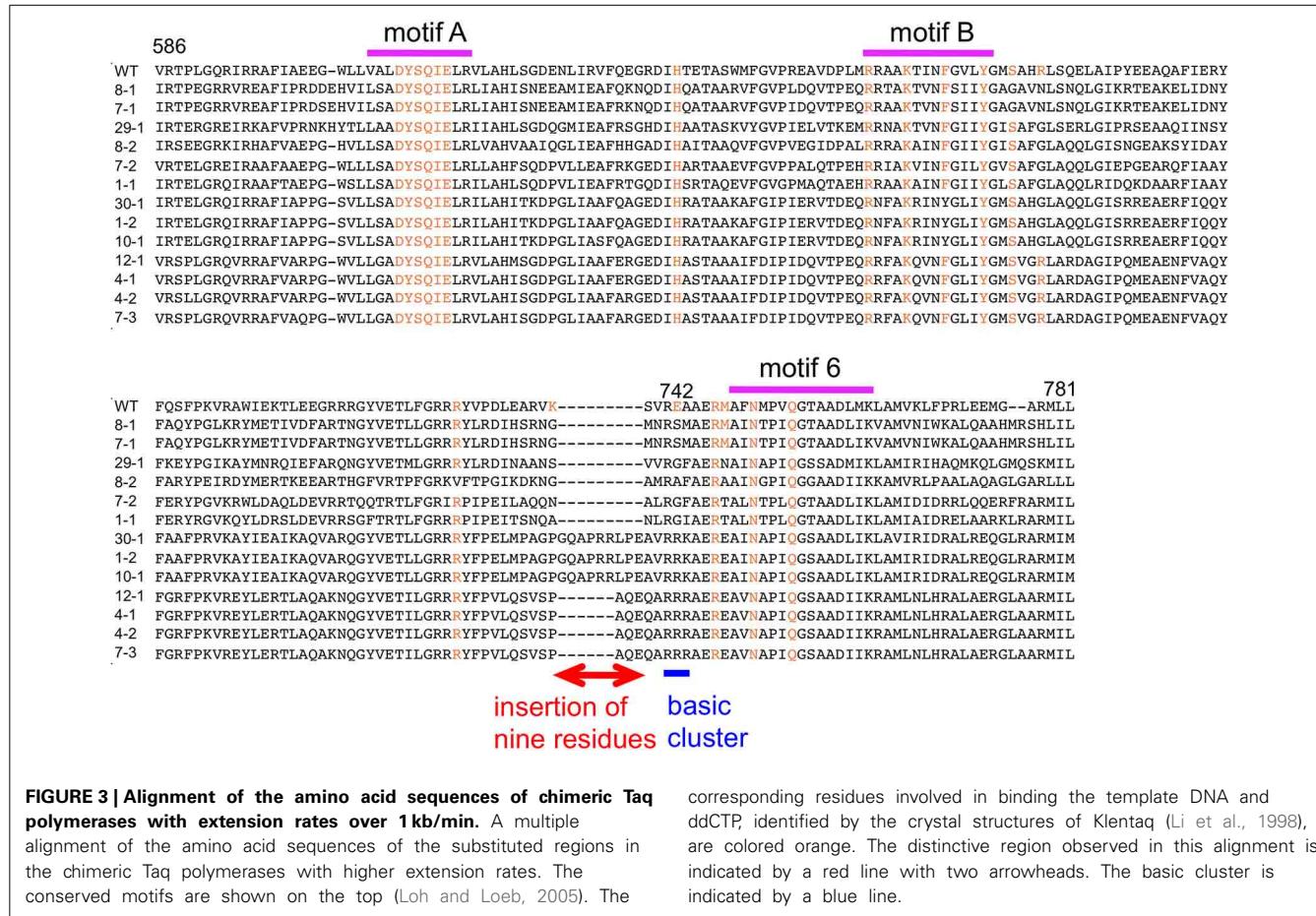


FIGURE 3 | Alignment of the amino acid sequences of chimeric Taq polymerases with extension rates over 1kb/min. A multiple alignment of the amino acid sequences of the substituted regions in the chimeric Taq polymerases with higher extension rates. The conserved motifs are shown on the top (Loh and Loeb, 2005). The

corresponding residues involved in binding the template DNA and ddCTP, identified by the crystal structures of Klentaq (Li et al., 1998), are colored orange. The distinctive region observed in this alignment is indicated by a red line with two arrowheads. The basic cluster is indicated by a blue line.

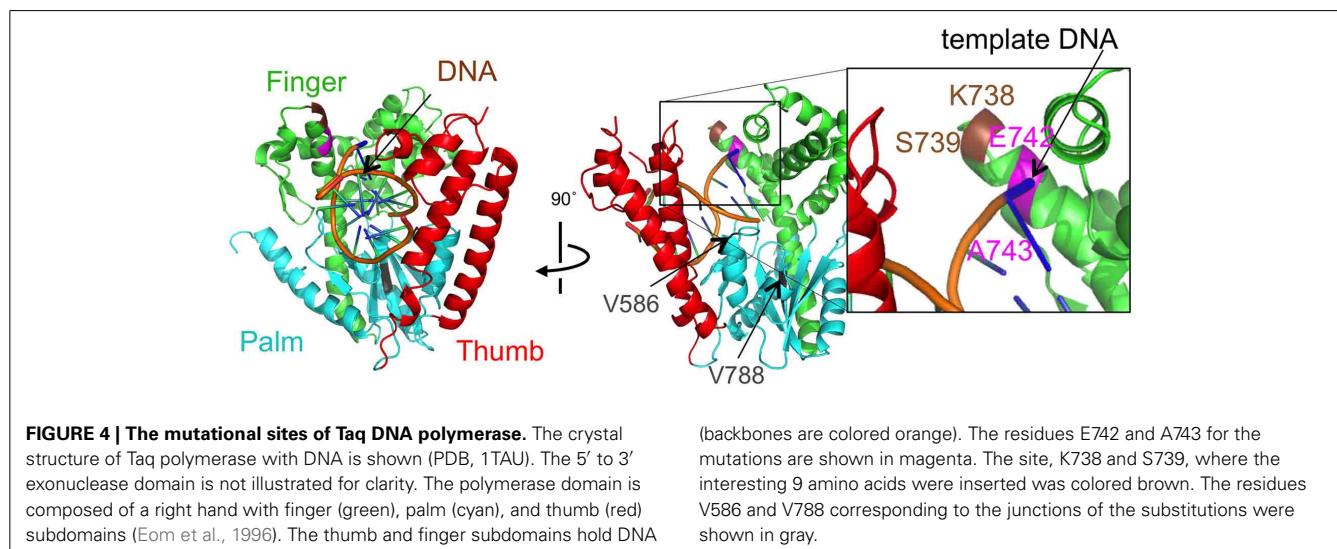


FIGURE 4 | The mutational sites of Taq DNA polymerase. The crystal structure of Taq polymerase with DNA is shown (PDB, 1TAU). The 5' to 3' exonuclease domain is not illustrated for clarity. The polymerase domain is composed of a right hand with finger (green), palm (cyan), and thumb (red) subdomains (Eom et al., 1996). The thumb and finger subdomains hold DNA

(backbones are colored orange). The residues E742 and A743 for the mutations are shown in magenta. The site, K738 and S739, where the interesting 9 amino acids were inserted was colored brown. The residues V586 and V788 corresponding to the junctions of the substitutions were shown in gray.

found at residues 741–743 in the chimeric Taq polymerases. The WT Taq polymerase has Glu and Ala at 742 and 743, but many chimeric enzymes showing faster extension have Arg at both positions. The crystal structure of the large fragment of Taq polymerase (Klentaq)-DNA complex revealed that Glu742

directly interacts with the template DNA in the closed conformation, but not in the open conformation (Li et al., 1998). As shown in the right panel of Figure 4, the residues Glu742 and Ala743 (magenta) are located in the finger subdomain and face to the template DNA (blue). The basic amino acid cluster in the

chimeric Taq polymerases is supposed to interact with the template DNA. We focused on this finding, and made a series of mutant polymerases by substitutions at positions 742 and 743 in WT Taq polymerase to change the affinity of the enzymes with DNA. The names of the mutant enzymes are as follows: RR (E742R/A743R), RA (E742R), AA (E742A), ER (A743R), AR (E742A/A743R), RK (E742R/A743K), KR (E742K/A743R), KK (E742K/A743K), QY (E742Q/A743Y), AH (E742A/A743H), EH (A743H), HA (E742H), HH (E742H/A743H), and HK (E742H/A743K). Fourteen mutant recombinant enzymes were purified to homogeneity from *E. coli* cells. The specific activity (units/mg protein) was measured by the standard incorporation assay (Table 4). Thermal stabilities of the mutant Taq polymerases were similar to WT enzyme (data not shown).

Table 4 | Properties of the mutant Taq polymerase.

Name	U/mg ($\times 10^5$)	Extention kb/5 min	Kd for DNA (nM)	Charge
TaqEA(WT)	5.1	0.7	400	-1
TaqRR	1.3	>8	9	2
TaqRA	2.3	3.7	8	1
TaqAA	2.8	2	9	0
TaqER	1.4	3.5	8	0
TaqAR	1.1	4.3	8	1
TaqRK	1.5	>8	10	2
TaqKR	1.1	>8	10	2
TaqKK	1.4	>8	10	2
TaqQY	4.7	2.1	10	0
TaqAH	3.7	2.5	6	1
TaqEH	3.2	2.0	6	0
TaqHA	3.8	2.5	6	1
TaqHH	3.8	2.8	6	2
TaqHK	2.2	4.8	6	2

FASTER PRIMER EXTENSION BY THE MUTANT Taq POLYMERASES

The *in vitro* primer extension rates were compared for these mutant Taq polymerases, as well as WT Taq polymerase. As shown in Figure 5, all of the mutant Taq polymerases exhibited faster extension reactions compared with that by the WT. The results of these experiments were quantified. The increased extension rate is generally related to the number of positive charges at this site (Table 4). The basic residues gave the varied effects. The positive charge of His appeared to have lower effect than those of Arg and Lys. There is a difference in pKa among the basic residues, and pKa of His, Arg and Lys are 6.8, 12.5, and 11.0, respectively. The relative degree of positive charge of His is estimated to be low. The DNA binding affinity of each enzyme was evaluated by EMSA, using a primed-DNA as a 32 P-labeled probe. As shown in Figure 6, the DNA binding ability of all the mutant Taq polymerases was distinctly increased from that of WT Taq polymerase. The increased number of positive charge at the positions 742 and 743 appeared to provide higher binding efficiency of Taq polymerase. Apparent Kd was determined with EMSA (Table 4). All the mutants bound to DNA by up to 2 orders of magnitude more tightly than WT Taq polymerase. Although there is no difference in the apparent Kd among these mutants, the second-shifted bands appeared in the gel images of EMSA in the case of the mutants, which possess Arg or Lys at the positions 742 and 743. The positive charge of Arg or Lys at the positions 742 and 743 might cause a nonspecific binding, in addition to the functional binding, of the enzyme to DNA.

BETTER PCR PERFORMANCE BY THE MUTANT Taq POLYMERASES

The main goal of this study was to create PCR enzymes with superior performance, as compared to that of WT Taq polymerase. Since the mutant Taq polymerases are as thermostable as WT enzyme, it was promising to apply these enzymes to PCR. Therefore, the PCR performances for several target DNAs with different lengths were compared. A representative example of the PCR experiments is shown in Figure 7. For the

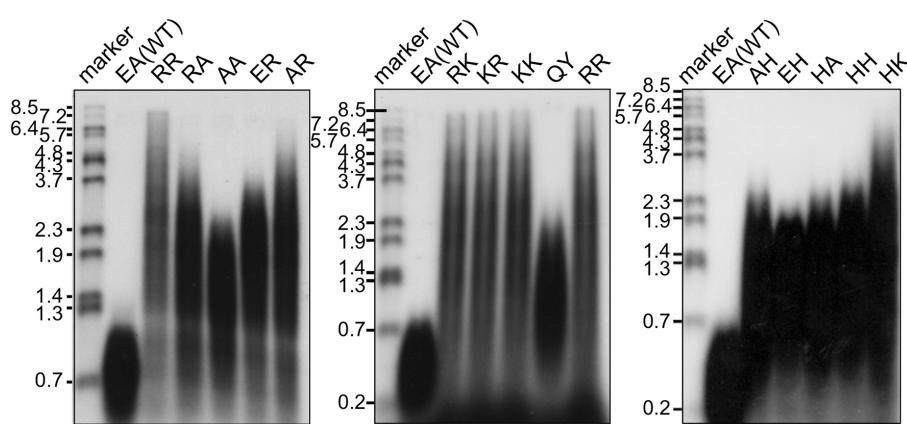


FIGURE 5 | Primer extension activities of WT and mutant Taq polymerases. M13 ssDNA annealed with a 32 P-labeled deoxyoligonucleotide (55mer) was used as the substrate. For each DNA polymerase, 0.05 unit was added to 20 μ l reaction mixture, containing 2.5 nM primed DNA. The reaction

mixtures were incubated at 74°C for 5 min, and the products were analyzed by 1% alkaline agarose gel electrophoresis, followed by autoradiography. The sizes indicated on the left are from BstPI-digested λ phage DNA labeled with 32 P at each 5' end. The names of the proteins were indicated on the top.

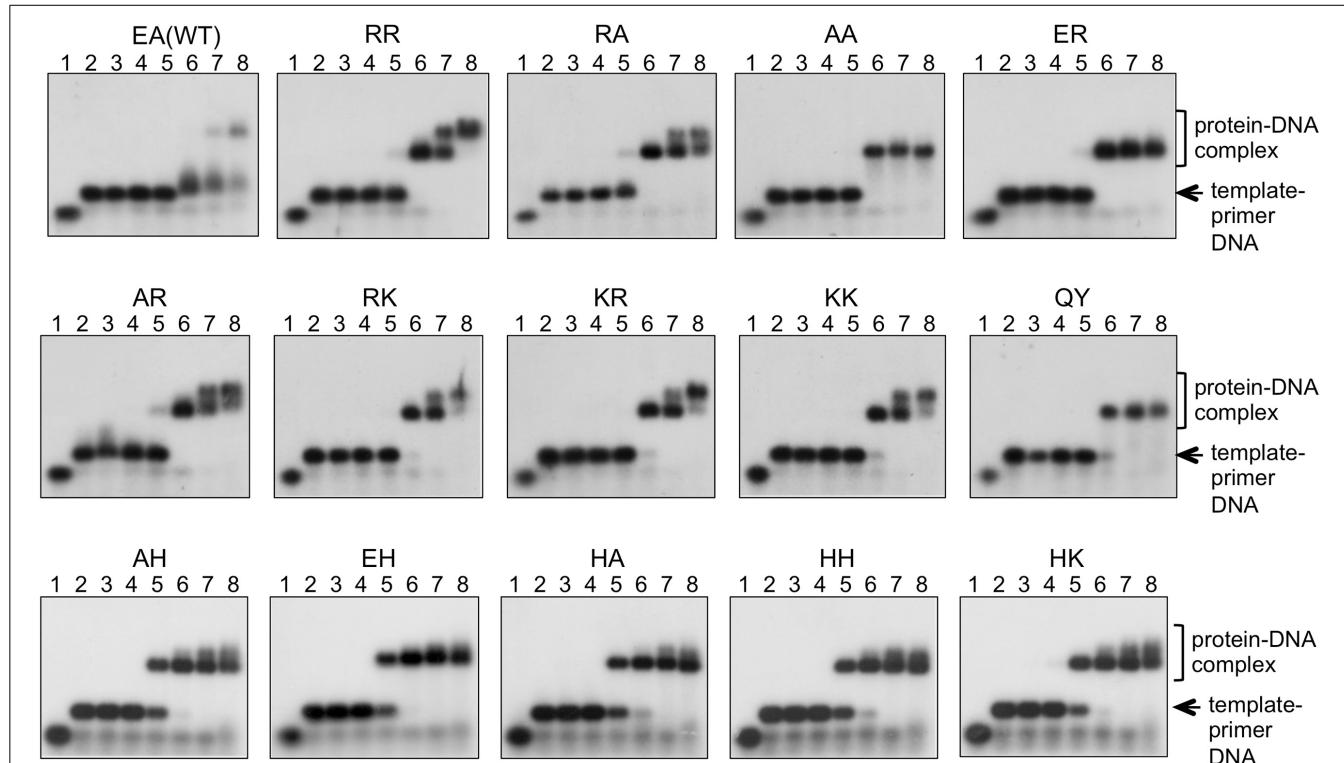


FIGURE 6 | DNA binding ability of mutant Taq polymerases. EMSA was performed using primed DNA (^{32}P -labeled 27mer ssDNA and 49mer DNA). The names of the mutant proteins were indicated on the top of each

panel. Lane 1 was ^{32}P -labeled 27mer ssDNA. Lane 2 had no protein with primed DNA. Lanes 3~8 contained 0.4, 1.6, 6.3, 25, 100, and 400 nM enzyme, respectively.

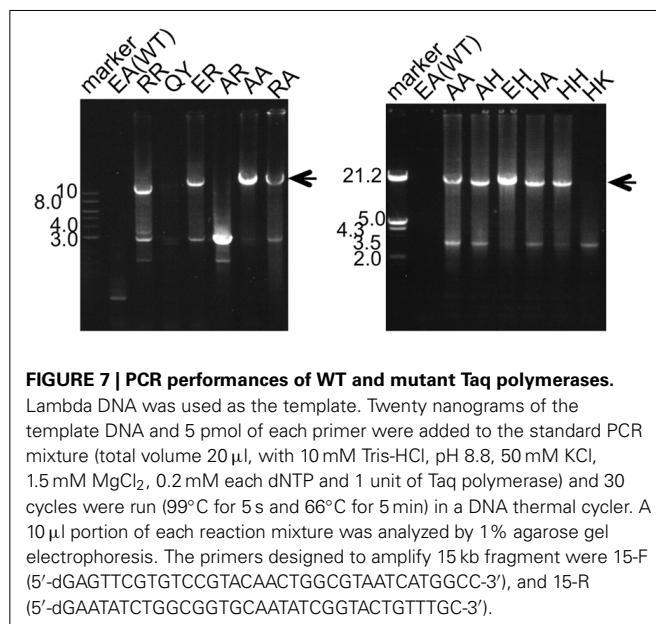


FIGURE 7 | PCR performances of WT and mutant Taq polymerases.

Lambda DNA was used as the template. Twenty nanograms of the template DNA and 5 pmol of each primer were added to the standard PCR mixture (total volume 20 μl , with 10 mM Tris-HCl, pH 8.8, 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM each dNTP and 1 unit of Taq polymerase) and 30 cycles were run (99°C for 5 s and 66°C for 5 min) in a DNA thermal cycler. A 10 μl portion of each reaction mixture was analyzed by 1% agarose gel electrophoresis. The primers designed to amplify 15 kb fragment were 15-F (5'-dGAGTCGTGTCCTACAACGGCGTAATCATGGCC-3'), and 15-R (5'-dGAATATCTGGCGGTCAATATCGGTACTGTTGC-3').

amplification of 15 kb of DNA, several mutant Taq polymerases, AA, RA, AH, EH, HA, and HH, successfully amplified the target DNA under conditions where WT Taq polymerase did not function. The other mutant enzymes prepared in this study did

not work well in the same conditions. The performances of some of the mutant enzymes, RR, QY, ER, AR, and HK, are shown in Figure 7. These experiments showed inconsistency with the results of primer extension experiment. The target DNA product was not detected by the mutant enzymes that possess primer extension rate with $>8 \text{ kb}/5 \text{ min}$. The enzymes with extension rate of $>8 \text{ kb}/5 \text{ min}$ have Arg or Lys at the positions 742 and 743. The observed PCR inhibition by these mutations may be due to the too tight binding to DNA, suggested by the EMSA as shown in the Section Faster Primer Extension by the Mutant Taq Polymerases. These results indicated that the positions of 742 and 743 in Taq polymerase are important for DNA strand synthesis, and the electrostatic environment of this site severely affects its PCR performance.

DISCUSSION

DNA polymerase is an important enzyme for both fundamental living phenomena (DNA replication/repair) in cells and applications to genetic engineering *in vitro*. Therefore, numerous structural and functional investigations of DNA polymerase have been reported to date. In this study, we developed PCR enzymes that provide a superior extension reaction as compared to Taq polymerase, the standard enzyme for PCR. As compared to the PCR performance of Taq polymerase, these enzymes achieved the amplification of either the same length of DNA in a shorter time or a longer DNA in the same reaction time.

Metagenomic analysis is a revolutionary technique for microbial ecology. The amplification of target genes from metagenomic DNA is a very powerful method to investigate many different DNA polymerases from uncultivated microbes. In this study, we focused on thermophilic bacteria as useful genetic resources for new thermostable family A DNA polymerases. We obtained many new sequences encoding a region of a family A DNA polymerase from the hot spring soil samples. These results suggested that our strategy to amplify a specific region of the family A DNA polymerase genes is actually applicable to the analysis of microbial populations in any habitat. We employed the same strategy to search for new family B DNA polymerases, and some of this work was published previously (Matsukawa et al., 2009).

We constructed chimeric enzymes between Taq polymerase and the products of the various *pol* genes amplified from the metagenomic DNAs, and their primer extension abilities were compared. Many chimeric polymerases possessing excellent extension ability were obtained by this experiment. However, none of the chimeric enzymes were sufficiently thermostable for PCR use. The microbial sources of the gene fragments used for the construction of the chimeric genes are not necessarily extreme thermophiles, and some moderate thermophiles and mesophiles may be included among the amplified genes. The chimeric Taq polymerases showing faster extension ability than WT Taq polymerase would have gene fragments from the organisms, which are not extreme thermophiles. However, the amino acid sequence comparison of the chimeric Taq polymerases provided an important clue to design a mutant Taq polymerase with superior speed for the primer extension reaction, by site-specific mutagenesis. We focused on positions 742 and 743 in this study. The positions 742 and 743 are located in the finger subdomain and affected the interaction with DNA. The PCR performances of some of the mutant Taq polymerases showed reliable improvement, and they are useful for faster PCR and also for longer target DNAs. The conversion of the electrostatic environment at this position, from a negative charge to a positive charge, will affect the stabilization of the DNA binding near the active site of Taq polymerase. It is important to check whether the mutations in this site affect the fidelity of Taq polymerase. Our preliminary data revealed that the fidelities of these enzymes are not different from that of WT Taq polymerase (data not shown). We will confirm this with more experiments and provide statistical data in the future.

In addition to positions 742 and 743, we found one more remarkable feature in the sequences of the chimeric enzymes. These enzymes have an insertion of either 9 or 3 amino acids between positions 738 and 739 of Taq polymerase. It will be interesting to investigate the effects of these insertions in the finger subdomain on the PCR performance of Taq polymerase. Characterizations of the mutant Taq polymerases with the different inserted sequences are now underway.

In conclusion, we designed a method for engineering Taq polymerase to improve its primer extension rate, by using information obtained from the metagenomic analysis of soil samples from various hot-spring areas. The created enzymes showed robust PCR performances that were better than that of Taq polymerase. Since Taq polymerase is the standard enzyme used for PCR, an abundance of PCR data using this enzyme has been accumulated

to date. The enzymes created in this study basically retain the properties of Taq polymerase, and therefore, they are applicable to many uses that have already been optimized with Taq polymerases.

ACKNOWLEDGMENTS

We thank Drs. Masaaki Takahashi and Yukiko Miyashita for valuable discussions and encouragement. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan [grant numbers 21113005, 23310152, and 26242075 to Yoshizumi Ishino]. This work was partly supported by Institute for fermentation, Osaka (IFO).

REFERENCES

- Baar, C., d'Abbadie, M., Vaisman, A., Arana, M. E., Hofreiter, M., Woodgate, R., et al. (2011). Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.* 39, e51. doi: 10.1093/nar/gkq1360
- Barns, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2216–2220. doi: 10.1073/pnas.91.6.2216
- Braithwaite, D. K., and Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21, 787–802. doi: 10.1093/nar/21.4.787
- Brakmann, S. (2005). Directed evolution as a tool for understanding and optimizing nucleic acid polymerase function. *Cell. Mol. Life Sci.* 62, 2634–2646. doi: 10.1007/s00018-005-5165-5
- Cann, I., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K., et al. (1999). Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J. Bacteriol.* 181, 6591–6599.
- Cann, I., and Ishino, Y. (1999). Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* 152, 1249–1267.
- d'Abbadie, M., Hofreiter, M., Caisman, A., Loakes, D., Gasparutto, D., Cadet, J., et al. (2007). Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* 25, 939–943. doi: 10.1038/nbt1321
- Eckert, K. A., and Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl.* 1, 17–24. doi: 10.1101/gr.1.1.17
- Eom, S. H., Wang, J., and Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature* 382, 278–281. doi: 10.1038/382278a0
- Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4552–4557. doi: 10.1073/pnas.071052198
- Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., et al. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* 22, 755–759. doi: 10.1038/nbt974
- Henry, A. A., and Romesberg, F. E. (2005). The evolution of DNA polymerases with novel activities. *Curr. Opin. Biotechnol.* 16, 370–377. doi: 10.1016/j.copbio.2005.06.008
- Holmberg, R. C., Henry, A. A., and Romesberg, F. E. (2005). Directed evolution of novel polymerases. *Biomol. Eng.* 22, 39–49. doi: 10.1016/j.bioeng.2004.12.001
- Ishino, S., and Ishino, Y. (2013). “DNA polymerases and DNA ligases,” in *Thermophilic Microbes in Environmental and Industrial Biotechnology*, eds J. Litterchild, T. Satyanarayana, and Y. Kawarabayasi (Dordrecht: Springer Science+Business Media), 429–457.
- Ishino, S., and Ishino, Y. (2014). DNA polymerase as a useful reagent for genetic engineering ~History of developmental research on DNA polymerase~. *Front. Microbiol.* 5:465. doi: 10.3389/fmicb.2014.00465
- Ishino, S., Kawamura, A., and Ishino, Y. (2012). Application of PCNA to processive PCR by reducing the stability of its ring structure. *J. Jap. Soc. Extremophiles* 11, 19–25.
- Ishino, Y., and Cann, I. (1998). The euryarchaeotes, a subdomain of archaea, survive on a single DNA polymerase: fact or farce? *Genes Genet. Syst.* 73, 323–336.
- Ishino, Y., Ueno, T., Miyagi, M., Uemori, T., Imamura, M., Tsunasawa, S., et al. (1994). Overproduction of *Thermus aquaticus* DNA polymerase and its structural analysis by ion-spray mass spectrometry. *J. Biochem.* 116, 1019–1024.

- Joyce, C. M., and Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63, 777–822. doi: 10.1146/annurev.bi.63.070194.004021
- Kermekchiev, M. B., Kirilova, L. I., Vail, E. E., and Barnes, W. M. (2009). Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Res.* 37, e40. doi: 10.1093/nar/gkn1055
- Kermekchiev, M. B., Tzakov, A., and Barnes, W. M. (2003). Cold-sensitive mutants of Taq DNA polymerase provide a hot start for PCR. *Nucleic Acids Res.* 31, 6139–6147. doi: 10.1093/nar/gkg813
- Komori, K., and Ishino, Y. (2000). Functional interdependence of DNA polymerizing and 3'→5' exonuclease activities in *Pyrococcus furiosus* DNA polymerase I. *Protein Eng.* 13, 41–47. doi: 10.1093/protein/13.1.41
- Li, Y., Kong, Y., Korolev, S., and Waksman, G. (1998). 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.* 17, 7514–7525. doi: 10.1093/emboj/17.24.7514
- Lipps, G., Röther, S., Hart, C., and Krauss, G. (2003). A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. *EMBO J.* 22, 2516–2525. doi: 10.1093/emboj/cdg246
- Loh, E., and Loeb, L. A. (2005). Mutability of DNA polymerase I: implications for the creation of mutant DNA polymerases. *DNA Repair* 4, 1390–1398. doi: 10.1016/j.dnarep.2005.09.006
- Matsukawa, H., Yamagami, T., Kawarabayasi, Y., Miyashita, Y., Takahashi, M., and Ishino, Y. (2009). A useful strategy to construct DNA polymerases with different properties by using genetic resources from environmental DNA. *Genes Genet. Syst.* 84, 3–13. doi: 10.1266/ggs.84.3
- Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., et al. (2001). The Y-family of DNA polymerases. *Mol. Cell* 8, 7–8. doi: 10.1016/S1097-2765(01)00278-7
- Ong, J. L., Loakes, D., Jaroslawski, S., Too, K., and Holliger, P. (2006). Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* 361, 537–550. doi: 10.1016/j.jmb.2006.06.050
- Perler, F. B., Kumar, S., and Kong, H. (1996). Thermostable DNA polymerases. *Adv. Protein Chem.* 48, 377–435. doi: 10.1016/S0065-3233(08)60367-8
- Suzuki, M., Avicola, A. K., Hood, L., and Loeb, L. A. (1997). Low fidelity mutants in the O-helix of *Thermus aquaticus* DNA polymerase I. *J. Biol. Chem.* 272, 11228–11235. doi: 10.1074/jbc.272.17.11228
- Suzuki, M., Yoshida, S., Adman, E. T., Blank, A., and Loeb, L. A. (2000). *Thermus aquaticus* DNA polymerase I mutants with altered fidelity. Interacting mutations in the O-helix. *J. Biol. Chem.* 275, 32728–32735. doi: 10.1074/jbc.M000097200
- Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., et al. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* 63, 4504–4510.
- Terpe, K. (2013). Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 97, 10243–10254. doi: 10.1007/s00253-013-5290-2
- Tosaka, A., Ogawa, M., Yoshida, S., and Suzuki, M. (2001). O-helix mutant T664P of *Thermus aquaticus* DNA polymerase I: altered catalytic properties for incorporation of incorrect nucleotides but not correct nucleotides. *J. Biol. Chem.* 276, 27562–27567. doi: 10.1074/jbc.M010635200
- Uemori, T., Ishino, Y., Doi, H., and Kato, I. (1995). The hyperthermophilic archaeon *Pyrodictium occultum* has two α-like DNA polymerases. *J. Bacteriol.* 177, 2164–2177.
- Uemori, T., Ishino, Y., Fujita, K., Asada, K., and Kato, I. (1993). Cloning of the *Bacillus caldothermophilus* DNA polymerase gene and characterization of the gene product. *J. Biochem.* 113, 401–410.
- Villbrandt, B., Sobek, H., Frey, B., and Schaumburg, D. (2000). Domain exchange: chimeras of *Thermus aquaticus* DNA polymerase, *Escherichia coli* DNA polymerase I and *Thermotoga neapolitana* DNA polymerase. *Protein Eng.* 13, 645–654. doi: 10.1093/protein/13.9.645
- Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* 32, 1197–1207. doi: 10.1093/nar/gkh271

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2014; accepted: 14 August 2014; published online: 03 September 2014.

Citation: Yamagami T, Ishino S, Kawarabayasi Y and Ishino Y (2014) Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering. *Front. Microbiol.* 5:461. doi: 10.3389/fmicb.2014.00461

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Yamagami, Ishino, Kawarabayasi and Ishino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota *Pyrococcus abyssi*

Melissa Castillo-Lizardo¹, Ghislaine Henneke^{2,3} and Enrique Viguera^{1*}

¹ Departamento de Biología Celular, Genética y Fisiología, Facultad de Ciencias, Universidad de Málaga, Málaga, Spain

² Laboratoire de Microbiologie des Environnements Extrêmes, UMR 6197, Institut Français de Recherche pour l'Exploitation de la Mer, Université de Bretagne Occidentale, Plouzané, France

³ CNRS, UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes, Plouzané, France

Edited by:

Zvi Kelman, University of Maryland, USA

Reviewed by:

Juergen Reichardt, James Cook University, Australia

Bernard Connolly, University of Newcastle, UK

***Correspondence:**

Enrique Viguera, Departamento de Biología Celular, Genética y Fisiología, Facultad de Ciencias, Universidad de Málaga, Campus Universitario de Teatinos, 29071, Málaga, Spain
e-mail: eviguera@uma.es

Replication slippage or slipped-strand mispairing involves the misalignment of DNA strands during the replication of repeated DNA sequences, and can lead to genetic rearrangements such as microsatellite instability. Here, we show that PolB and PolD replicative DNA polymerases from the archaeal model *Pyrococcus abyssi* (*Pab*) slip *in vitro* during replication of a single-stranded DNA template carrying a hairpin structure and short direct repeats. We find that this occurs in both their wild-type (exo+) and exonuclease deficient (exo-) forms. The slippage behavior of *Pab*PolB and *Pab*PolD, probably due to limited strand displacement activity, resembles that observed for the high fidelity *P. furiosus* (*Pfu*) DNA polymerase. The presence of *Pab*PCNA inhibited *Pab*PolB and *Pab*PolD slippage. We propose a model whereby *Pab*PCNA stimulates strand displacement activity and polymerase progression through the hairpin, thus permitting the error-free replication of repetitive sequences.

Keywords: slippage, primer-template misalignment, DNA polymerases, strand displacement activity, Archaea

INTRODUCTION

Low complexity DNA sequences such as microsatellites (1–9 nt repeat length), including mono, di, and trinucleotide repeats, and minisatellites (unit ≥ 10 nt) are frequently associated with mutagenesis “hot-spots” in both eukaryotic and prokaryotic genomes (Bierne et al., 1991; Michel, 2000; Aguilera and Gomez-Gonzalez, 2008). These types of sequences are characterized by high instability, consisting of the addition or deletion of repeated units, leading to variations in repeat copy number. Such genetic variations have been termed “dynamic mutations” (Richards and Sutherland, 1992; Pearson et al., 2005). Arrest of the replication machinery within a repeated region is associated with such instability, where primer and template become misaligned (reviewed in Michel, 2000). This process, known as replication slippage, is involved in the generation of deletions or insertions within repeat regions (Viguera et al., 2001a; Lovett, 2004).

Replication slippage has been proposed to occur within homopolymeric runs (Kroutil et al., 1996) as well as in short and long tandem repeat sequences (Trinh and Sinden, 1993; Madsen et al., 1993; Tran et al., 1995; Bierne et al., 1996; Feschenko and Lovett, 1998). Repeated DNA sequences are generally characterized by the formation of non-B DNA structures, the majority of which can form intra-strand hairpin loops (Samadashwily et al., 1997; McMurray, 1999; Mirkin, 2007; Sinden et al., 2007). A direct role for replication slippage in the deletion of repeated sequences within hairpin structures has been demonstrated *in vitro* and *in vivo* (d’Alençon et al., 1994; Caneill and Ehrlich, 1996). Slippage-mediated deletions are believed to occur via a three step mechanism as illustrated in Figure 1 (Viguera et al., 2001a). In this model, the polymerase pauses as it reaches the base of the

hairpin after copying the first direct repeat (DR), followed by polymerase dissociation. The 3' end of the nascent strand then unpairs from the template before reannealing to the second DR. This new primer/template complex is recognized by the polymerase, allowing replication to continue but also generating a deletion.

Several DNA polymerases have been tested for their propensity to slip *in vitro* when replicating hairpin-containing templates. Surprisingly, the replicative DNA polymerase Pol III holoenzyme (HE) from *Escherichia coli* can slip *in vitro* (Caneill and Ehrlich, 1996; Caneill et al., 1999). This is of utmost importance because high fidelity replication is required to maintain genome integrity. Studies on DNA polymerases involved in DNA repair such as *E. coli* Pol I, *E. coli* Pol II, and the T4, T7, and ϕ 29 phage DNA polymerases revealed that the strand displacement activity of a DNA polymerase is inversely related to their propensity to slip. DNA polymerases with high strand displacement activity such as ϕ 29 or T7 pol exo- (SequenaseTM) can progress through template hairpin structures and consequently do not slip. On the other hand, DNA polymerases devoid of strand displacement activity such as *E. coli* DNA Pol II or T4 are blocked at the base of the hairpin, promoting DNA repeat misalignment and subsequent loss of repeat sequences. Depending on the template and the strand displacement activity of a DNA polymerase it is possible for the deletion error rate to exceed the base substitution rate (Kunkel and Bebenek, 2000). In the context of the model proposed for slippage-mediated deletions (Figure 1), a polymerase with high strand displacement activity would be able to open the hairpin duplex, avoiding the polymerase dissociation and nascent strand reannealing steps, and thus replicate the repeat-containing template faithfully.

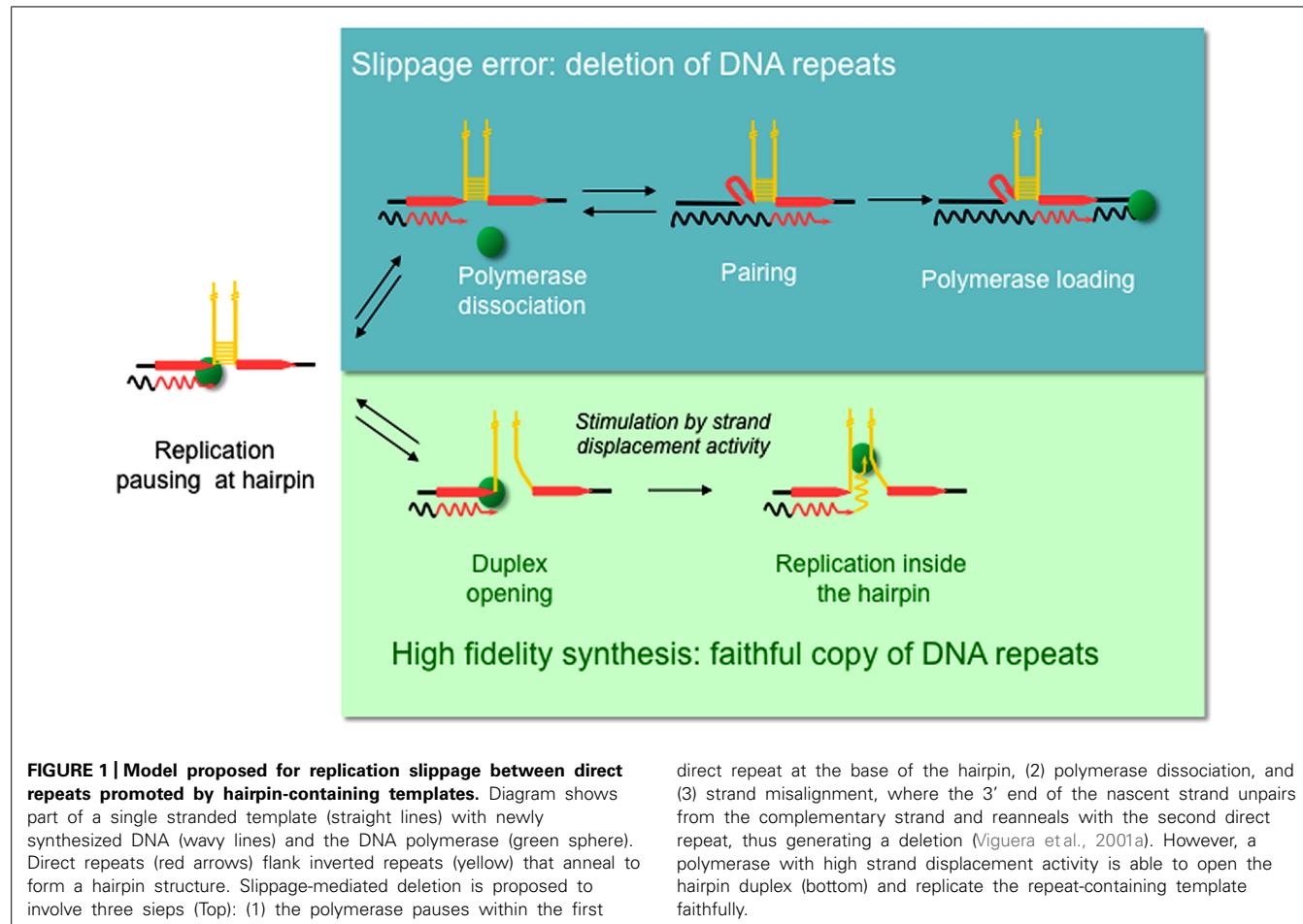


FIGURE 1 | Model proposed for replication slippage between direct repeats promoted by hairpin-containing templates. Diagram shows part of a single stranded template (straight lines) with newly synthesized DNA (wavy lines) and the DNA polymerase (green sphere). Direct repeats (red arrows) flank inverted repeats (yellow) that anneal to form a hairpin structure. Slippage-mediated deletion is proposed to involve three steps (Top): (1) the polymerase pauses within the first

direct repeat at the base of the hairpin, (2) polymerase dissociation, and (3) strand misalignment, where the 3' end of the nascent strand unpairs from the complementary strand and reanneals with the second direct repeat, thus generating a deletion (Viguera et al., 2001a). However, a polymerase with high strand displacement activity is able to open the hairpin duplex (bottom) and replicate the repeat-containing template faithfully.

Several thermostable DNA polymerases utilized for PCR can also slip even under the high temperatures used during PCR amplification (Viguera et al., 2001b); **Table 1.** DNA polymerases with high fidelity in terms of base substitution rate, such as *P. furiosus* (*Pfu*Pol), consistently undergo slippage and introduce deletions while replicating hairpin-containing templates (Viguera et al., 2001b). In contrast, a low fidelity DNA polymerase such as *Thermus aquaticus* (*Taq*Pol) can replicate the same hairpin sequence without introducing deletions, although this is dependent on the magnesium concentration used. Other thermostable DNA polymerases endowed with a high strand displacement activity such as *Thermococcus fumicola* (*Tfu*Pol) or *Bacillus stearothermophilus* (*Bst*Pol) also do not slip when replicating hairpin-containing sequences (Viguera et al., 2001b).

We have studied here the biochemical properties of DNA polymerases PolB (*Pab*PolB) and PolD (*Pab*PolD) from the hyperthermophilic euryarchaeon *P. abyssi* in terms of slippage during *in vitro* primer extension reactions. Archaeal replication proteins are more closely related to their eukaryotic than their bacterial equivalents. Euryarchaeal members contain DNA polymerases that belong to both the ubiquitous B family as well as the D family, which is unique to archaea (Ishino et al., 1998; Barry and Bell, 2006; Raymann et al., 2014). Both PolB and PolD have associated 3'-5' exonuclease activity and moderate strand

displacement activity, although PolB cannot displace a RNA-DNA hybrid (Henneke, 2012). However, in the presence of *Pab* Proliferating cell nuclear antigen (PCNA), both *Pab* polymerases show strand displacement activity (Henneke et al., 2005; Rouillon et al., 2007). Moreover, *Pab*PCNA can be loaded onto DNA in the absence of the clamp-loader replication factor C (RF-C), although the presence of this factor does enhance its loading (Rouillon et al., 2007).

In this work, we report that both *P. abyssi* DNA polymerases slip *in vitro* on a template that consists of single-stranded DNA (ssDNA) with a hairpin structure flanked by short direct repeats. In addition, we find that *Pab*PCNA increases replication fidelity of this template by triggering the strand displacement activity of *Pab*polB. Furthermore, we describe the effect of magnesium concentration on the replication slippage of both *Pab* DNA polymerases. These results help toward understanding the dynamics of replication through common non-B DNA structures and identifying the key DNA polymerases involved in replication slippage; a crucial step for understanding genome stability in these organisms.

MATERIALS AND METHODS

PROTEINS

*Pab*PCNA, *Pab*pol D, and exonuclease-deficient *Pab*pol D were obtained from G. Henneke (Ifremer, Brest, France). They were

Table 1 | Effect of magnesium concentration on the slippage of *Pfu*, *Taq*, *Vent* (50°C), *Vent* (65°C), *Tfu* and *Bst* polymerases determined previously (Viguera et al., 2001b) and *Pab PolB*, *Pab PolB exo-*, *Pab PolD*, and *Pab PolD exo-* determined in this work.

Polymerase	Magnesium concentration (mM)							
	0.5	1	2.5	5	7.5	10	15	20
<i>Pfu Pol</i>	S	S	S	S	S	-	-	nd
<i>Pfu Pol</i> (native)	-	P/S	P/S	S	S	-	-	-
<i>Taq Pol</i>	P	P	P	P/S	P/S	S	S	-
<i>Vent Pol</i> (50°C)	-	-	S	S	S	S	-	nd
<i>Vent Pol</i> (65°C)	-	-	rcr	rcr	P/S	P/S	-	nd
<i>Tfu Pol</i>	nd	rcr	nd	nd	nd	nd	nd	nd
<i>Bst Pol</i>	-	-	rcr	rcr	rcr	rcr	nd	nd
<i>Pab PolB</i>	-	P/S	P/S	P/S	P/S	S	S	-
<i>Pab PolB exo-</i>	-	P/S	P/S	P/S	S	S	-	-
<i>Pab PolD</i>	-	S	P/S	P/S	S	S	S	-
<i>Pab PolD exo-</i>	-	P/S	P/S	P/S	S	S	S	S

The main product obtained for each reaction is shown. S indicates slipped molecules, generated by replication slippage error. P indicates parental molecules, indicative of faithful replication. rcr indicates high molecular weight molecules generated by rolling circle replication as a consequence of the strand displacement activity of a DNA polymerase. nd, not determined.

cloned, expressed and purified as described (Gueguen et al., 2001; Henneke et al., 2002, 2005; Palud et al., 2008). *PabpolB* (Isis™) and *PabpolB* exonuclease-deficient (*Pyra*™ exo-) were purchased from MP Biomedicals. One unit of *Pab* pols corresponds to the incorporation of 1 nmol of total dTMP into acid precipitable material per minute at 65°C in a standard assay containing 0.5 µg (nucleotides) of poly(dA)/oligo(dT)_{10:1} M13 gene protein II (gp II) was purified to homogeneity as described (Greenstein and Horiuchi, 1987). *Thermus thermophilus* SSB was a kind gift from Drs. C. Perales and J. Berenguer (CBM-SO, Madrid). Native *Pfu* Pol was from Stratagene. *Taq* Pol was from Roche Molecular Biochemicals.

ssDNA TEMPLATE

Construction of the pHP727FXc plasmid has been described previously (Canceill and Ehrlich, 1996). Preparation of ssDNA templates was carried out essentially as described (Canceill and Ehrlich, 1996) with the following modifications: plasmid DNA was extracted using a Maxi Plasmid Kit (Qiagen). Briefly, a specific nick was introduced into the f1 replication origin (+) strand of purified FXc plasmid DNA using the M13 gpII protein. The reaction was stopped with 20 mM EDTA and the products treated with 200 µg/ml proteinase K for 10 min at 55°C, phenol extracted and dialyzed against TE buffer. Nicked strands were removed by exonuclease III digestion (10–40 units per µg of DNA for 1 h at 37°C). Finally, Exo III, nucleotides and oligonucleotides were removed using QIAquick® PCR (Qiagen) purification kits.

PRIMER EXTENSION REACTIONS

Pyrococcus abyssi pols were tested in a primer extension reaction performed as described (Canceill and Ehrlich, 1996; Canceill et al.,

1999). Briefly, 24.3 fmol of a 5'-end fluorescein labeled primer (Applied Biosystems) designated 1233 (5'AGC GGA TAA CAA TTT CAC ACA GGA 3'), were annealed 1235 bases upstream of the palindrome. All 10 µl primer extension reactions contained 25 ng (12.2 fmol) of primed ssDNA, and *Pab* pols that were added to the reaction mixture as indicated in the figure legends. Additionally, reactions contained unless otherwise mentioned 50 mM Tris-HCl (pH 8.8), 50 mM KCl, 10 mM DTT, 2 mM MgCl₂, and 200 µM dNTPs. Reactions were performed at 60°C for 30 min, synthesis was arrested by the addition of 25 mM EDTA and 500 µg/ml proteinase K, and the mixture was further incubated for 15 min at 55°C. Reaction products were analyzed by electrophoresis using 0.8% agarose gels under native conditions, run in TBE buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.3) at 2 V/cm for 16 h and visualized with a Typhoon 9400 Variable Mode Imager (Amersham Biosciences, GE Healthcare). Analysis of the results was performed using Image Quant 5.2 software. Quantification analysis was performed with Visilog 6.3 (Visualization Sciences Group. Noesis). A common fixed area was selected at the center of the bands corresponding to parental and heteroduplex molecules. The average gray value of all pixels of each area was obtained and the proportion of parental/heteroduplex was calculated.

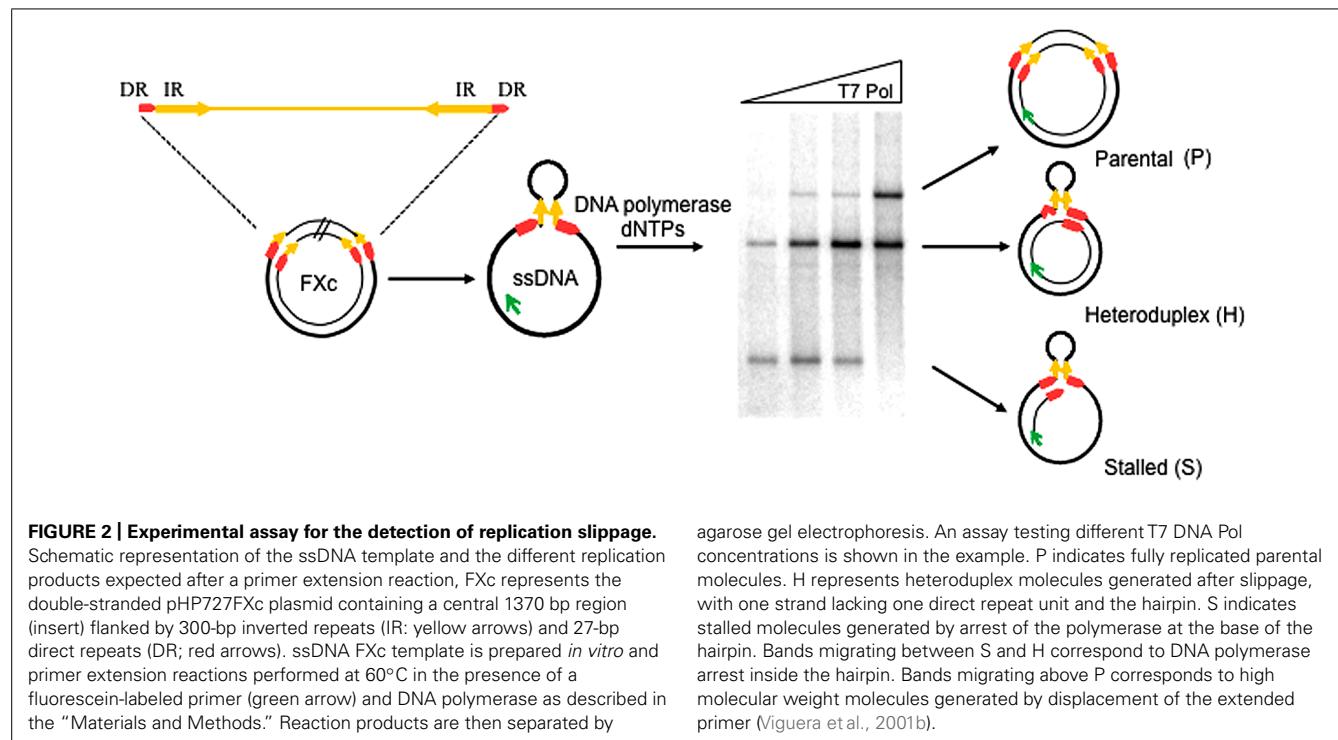
Pfu Pol and *Taq* Pol were tested as above except that 200 µM dGTP, dATP, and dTTP (each), 40 µM dCTP and 50 µM (2.5 µCi) (α -³²P)dCTP was used. The reaction buffers were prepared magnesium free as those furnished by the suppliers and contained, in addition to 30 mM NaCl brought by the primed ssDNA, the following ingredients: (i) for *Taq* Pol: 10 mM Tris-HCl pH 8.3, 50 mM KCl; (ii) for Native *Pfu* Pol: 20 mM Tris-HCl pH 8.0, 10 mM KCl, 6 mM (NH₄)₂SO₄, 0.1% Triton® X-100, 10 µg/ml BSA. After gel electrophoresis, DNA was visualized by direct exposure of the dried gels to Imaging Plates (IP BAS-MP 2040S) and analyzed on a Fujifilm-BAS 1500.

RESULTS

EXPERIMENTAL SYSTEM

To study whether *P. abyssi* thermostable DNA polymerases (*Pab* pols) promote replication slippage, we performed primer-extension assays using the circular ssDNA template, FXc (Canceill and Ehrlich, 1996; Canceill et al., 1999; Viguera et al., 2001a,b). This template contains two 27 bp direct repeats (DRs) that flank a pair of 300 bp inverted repeats (IR) separated by a 1.3 kb insert as shown in Figure 2. The IRs anneal to form a stem-loop with the DRs at its base.

DNA synthesis was carried out with a fluorescently labeled primer and the reaction products analyzed by agarose gel electrophoresis. Faithful replication of the FXc template generates complete double-stranded parental (P) molecules, which migrate in a retarded position on the gel. A slippage event generates a heteroduplex molecule (H), composed of a parental strand annealed to a recombinant strand lacking one of the DRs and the 1370 bp region between them. Heteroduplex molecules migrate ahead of parental molecules. Stalled (S) replication as the polymerase reaches the base of the hairpin results in a truncated molecule that migrates further than either parental or heteroduplex molecules.



SLIPPAGE OF *P. abyssi* PolB AND PolD REPLICATIVE POLYMERASES

It has been proposed that PolB and PolD have different roles in the cell, both participating at the replication fork in a manner analogous to *Bacillus subtilis* and the eukaryotic replisome. The current model for *P. abyssi* DNA replication proposes that *PabPolD* performs RNA-primed DNA synthesis and is later displaced by *PabPolB* to carry out processive DNA synthesis, at least on the leading strand (Henneke et al., 2005; Rouillon et al., 2007). Because of the capacity of *PabPolD* to displace RNA primers in a PCNA-dependent manner, it has been suggested that *PabPolD* is involved in lagging strand replication. However, definitive confirmation using genetic approaches such as those employed with eukaryotic polymerases has yet to be performed.

In order to test whether their putatively separate roles in leading and lagging strand replication also imply different slippage properties, we examined the slippage efficiency of wildtype *PabPolB* and *PabPolD* enzymes using the FXc template (Figure 2). *PabPolB* generated both parental and heteroduplex molecules, which indicate a mixture of normal FXc replication and slippage events (Figure 3, lanes 1–3). Similar proportions of parental and heteroduplex molecules were produced by *PabPolB*, with a slightly higher ratio of parental molecules as the polymerase concentration was increased. *PabPolD* behaved in a similar way although the proportion of heteroduplex molecules was higher and overall synthesis was improved at higher polymerase concentrations (Figure 3, lanes 7–9). These results indicate that *PabPolB* and *PabPolD* can slip under our assay conditions. The behavior of *PabPolB* and *PabPolD* is similar to that observed for Pol III HE, T7 Pol, or *Taq* Pol that also produce both parental and heteroduplex molecules (Canceill and Ehrlich, 1996; Canceill et al., 1999; Viguera et al., 2001b).

To generate parental molecules, a DNA polymerase must open the hairpin formed by the annealed inverted repeats of the single-stranded template (Figure 2), which is largely dependent on a DNA polymerase's strand displacement activity. As a consequence, polymerases with high strand displacement activity (e.g., ϕ 29 DNA polymerase) do not slip while DNA polymerases devoid of strand displacement activity (e.g., *E. coli* Pol II or T4 DNA pol) generate heteroduplex molecules as the sole product of the reaction (Canceill et al., 1999).

Strand displacement activity is modified in some DNA polymerase exo- mutants. For example the T7 DNA polymerase has relatively low strand displacement activity (Canceill et al., 1999). However, the T7 pol exo- (SequenaseTM), carrying a 28 amino acid deletion that inactivates its proofreading activity (Engler et al., 1983; Lechner et al., 1983), has increased strand displacement activity that prevents T7 pol exo- slippage (Canceill et al., 1999). Similarly, the *E. coli* Pol II exo- mutant gains a degree of strand displacement activity and the ability to synthesize parental molecules (Canceill et al., 1999). However, not all exo- forms exhibit increased strand displacement activity. For example, an exo- form of F 29 caused by a point mutation shows a 90% reduction in strand displacement activity compared to the native enzyme (Soengas et al., 1992).

To test whether *Pab* pol exo- variants have modified slippage properties, we performed FXc template primer extension assays using exo- mutant forms of *PabPolB* and *PabPolD* carrying single point mutations (D215A and H451A, respectively; see Palud et al., 2008). Our results show that the exo- forms of *PabPolB* and *PabPolD* both generated heteroduplex and parental molecules. However, unlike their native forms, increasing polymerase concentration inhibited slippage and resulted in a higher

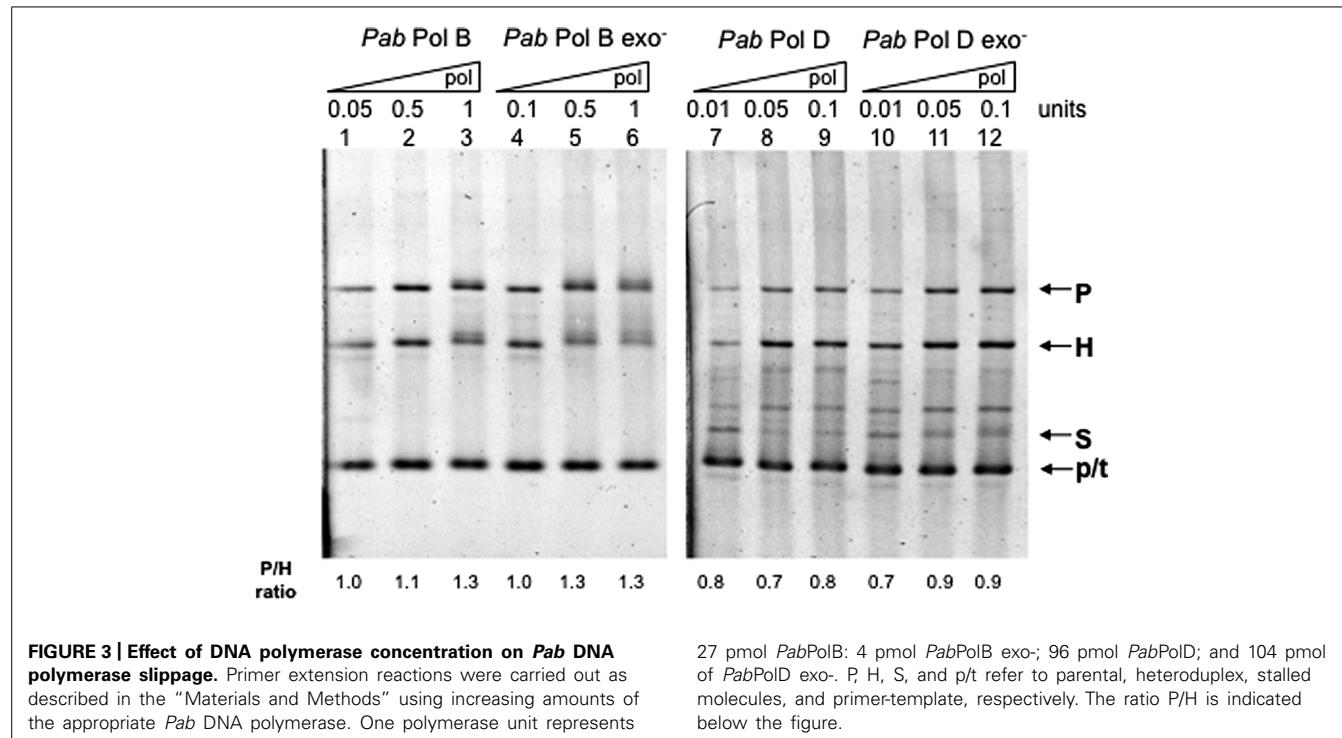


FIGURE 3 | Effect of DNA polymerase concentration on *Pab* DNA polymerase slippage. Primer extension reactions were carried out as described in the “Materials and Methods” using increasing amounts of the appropriate *Pab* DNA polymerase. One polymerase unit represents

27 pmol *Pab*PolB; 4 pmol *Pab*PolB exo-; 96 pmol *Pab*PolD; and 104 pmol of *Pab*PolD exo-. P, H, S, and p/t refer to parental, heteroduplex, stalled molecules, and primer/template, respectively. The ratio P/H is indicated below the figure.

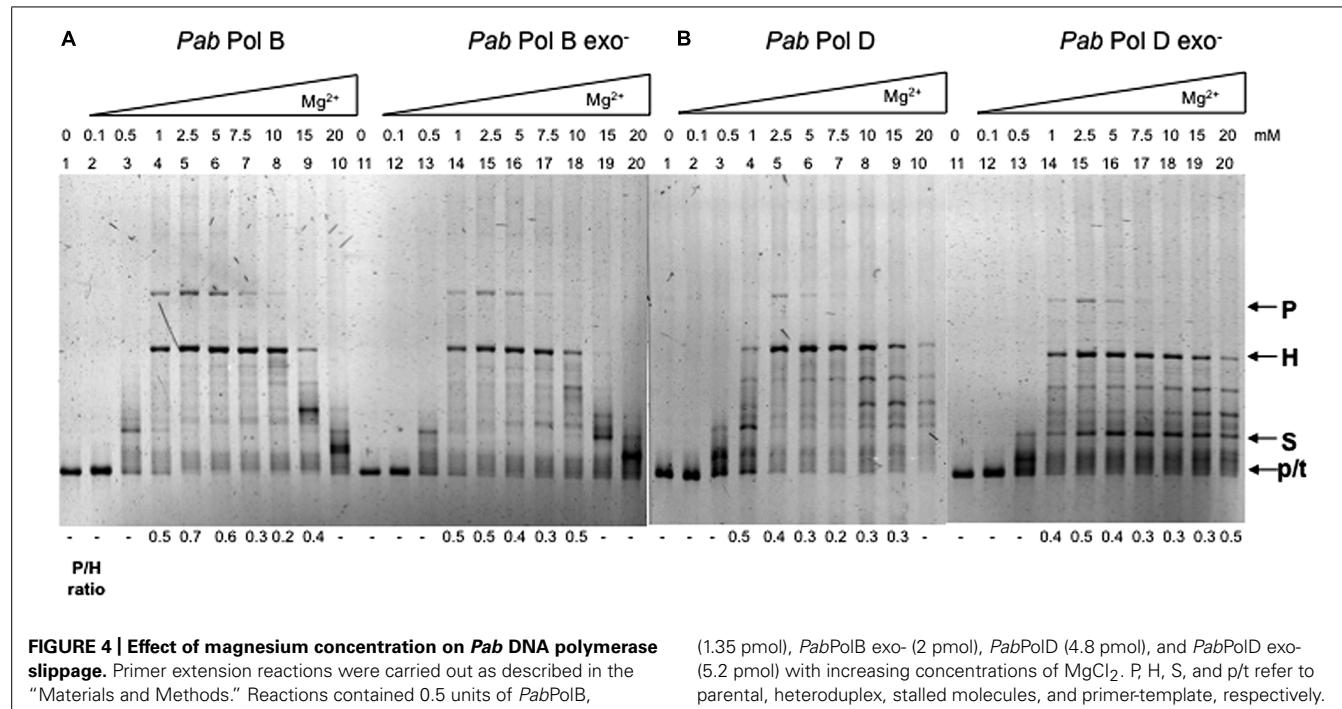
proportion of parental molecules (Figure 3, lanes 4–6 and 10–12, respectively). Both the native *Pab*PolD enzyme and its exo- form produced some molecules that migrate between the heteroduplex and stalled molecules, possibly the result of inefficient polymerase progression through the hairpin (Viguera et al., 2001a).

MAGNESIUM CONCENTRATION AFFECTS THE SLIPPAGE OF *P. abyssi* POLYMERASES

The concentration of divalent cations needs to be precisely controlled during DNA synthesis as it affects enzyme activity, enzyme fidelity, primer/template annealing, and the stability of secondary structures, such as the stem-loop used in our assay. The fidelity of *Taq* and *Pfu* DNA polymerases in terms of base substitution and frameshift errors is dependent on magnesium concentration (Eckert and Kunkel, 1990; Cline et al., 1996). Moreover, trinucleotide repeat expansions are produced *in vitro* by *Taq*, *E. coli* Pol I, and the Pol I Klenow fragment at certain magnesium concentrations (Lyons-Darden and Topal, 1999). With respect to slippage, magnesium concentration differentially affects the slippage errors produced by thermostable DNA polymerases (Viguera et al., 2001b). *In vitro* experiments showed that slippage error-derived heteroduplex molecules account for almost all the product generated by *Pfu* DNA polymerase over the magnesium concentration range that permits efficient DNA synthesis (0.5–7.5 mM MgSO₄). In contrast, from the same template *Taq* DNA polymerase faithfully generates parental molecules at a low magnesium concentration (0.5 mM MgCl₂), heteroduplex molecules at high magnesium concentrations (10–20 mM MgCl₂) and a mixture of parental and heteroduplex molecules at intermediate magnesium concentrations (1–7.5 mM MgCl₂; Viguera et al., 2001b).

These observations prompted us to analyze the effect of magnesium concentration on the slippage errors produced by the wildtype and exo- forms of *Pab*PolB and *Pab*PolD. We found that varying magnesium concentration affected both slippage and overall DNA synthesis (Figure 4). There was almost no synthesis by *Pab*PolB, *Pab*PolD or their exo- forms at low magnesium concentrations (0.1–0.5 mM; Figure 4A, lanes 2–3 and 12–13; Figure 4B, lanes 2–3 and 12–13). Synthesis was also inhibited at the highest concentrations tested (15–20 mM; Figure 4A, lanes 9–10 and 19–20; Figure 4B, lanes 9–10 and 19–20). Parental molecules were readily detectable together with heteroduplex molecules at low to medium magnesium concentrations (1–5 mM; Figure 4A, lanes 4–6). Increasing magnesium concentration up to 15 mM decreased the proportion of parental molecules and resulted in heteroduplex molecules as the main reaction product (Figure 4A, lanes 7–9). This latter result could be due to stabilization of the hairpin structure by high magnesium concentrations making polymerase progression more difficult inside the hairpin (Cancell and Ehrlich, 1996). The effect of magnesium concentration on *Pab*PolB was similar at the 40 and 100 μM nucleotide concentrations tested (data not shown). *Pab*PolB exo-behaved in a similar way to the wildtype enzyme (Figure 4A, lanes 14–18).

Although both *Pab*PolD and *Pab*PolD exo- generated parental molecules, the main reaction products were heteroduplex molecules whenever synthesis was efficient (Figure 4B, lanes 4–10 and lanes 14–20). Additionally, some of the molecules generated by *Pab*PolD and *Pab*PolD exo- migrated between heteroduplex and stalled molecules (Figure 4B, lanes 8–9 and 15–20) that probably represent partially replicated DNA molecules due to inefficient polymerase progression within the hairpin.



In order to confirm that the magnesium concentrations used in the previous experiment are compatible with efficient *Pab*PolB and *Pab*PolD DNA synthesis, we performed primer extension experiments using a 5' fluorescently labelled primer (33 mer) and a short single-stranded linear DNA template (87 mer) that has the potential to form a 28 bp secondary structure but lacks DRs (Henneke, 2012). This assay should only assess the replication efficiency of a template with a small hairpin without the possibility of slippage between DRs at its base. Fully replicated molecules were generated for both *Pab*PolB and *Pab*PolD using the same range of magnesium concentrations used for the primer extension assays (Figure S1), which indicates that the reaction conditions used were optimal for DNA synthesis.

We conclude that in spite of their high fidelity in terms of base substitution, *Pab*PolB and *Pab*PolD are highly prone to slip on ssDNA templates upon encountering secondary structures flanked by DRs, generating parental and heteroduplex molecules in a magnesium concentration-dependent manner. This is in agreement with previous results describing similar behavior for *Taq* DNA polymerase (Viguera et al., 2001b); Figure S2. In contrast, native *Pfu* DNA polymerase generates mostly heteroduplex molecules regardless of magnesium concentration (Viguera et al., 2001b); Figure S2.

Pyrococcus abyssi PCNA CAN MODULATE THE SLIPPAGE OF *Pab*PolB AND *Pab*PolD

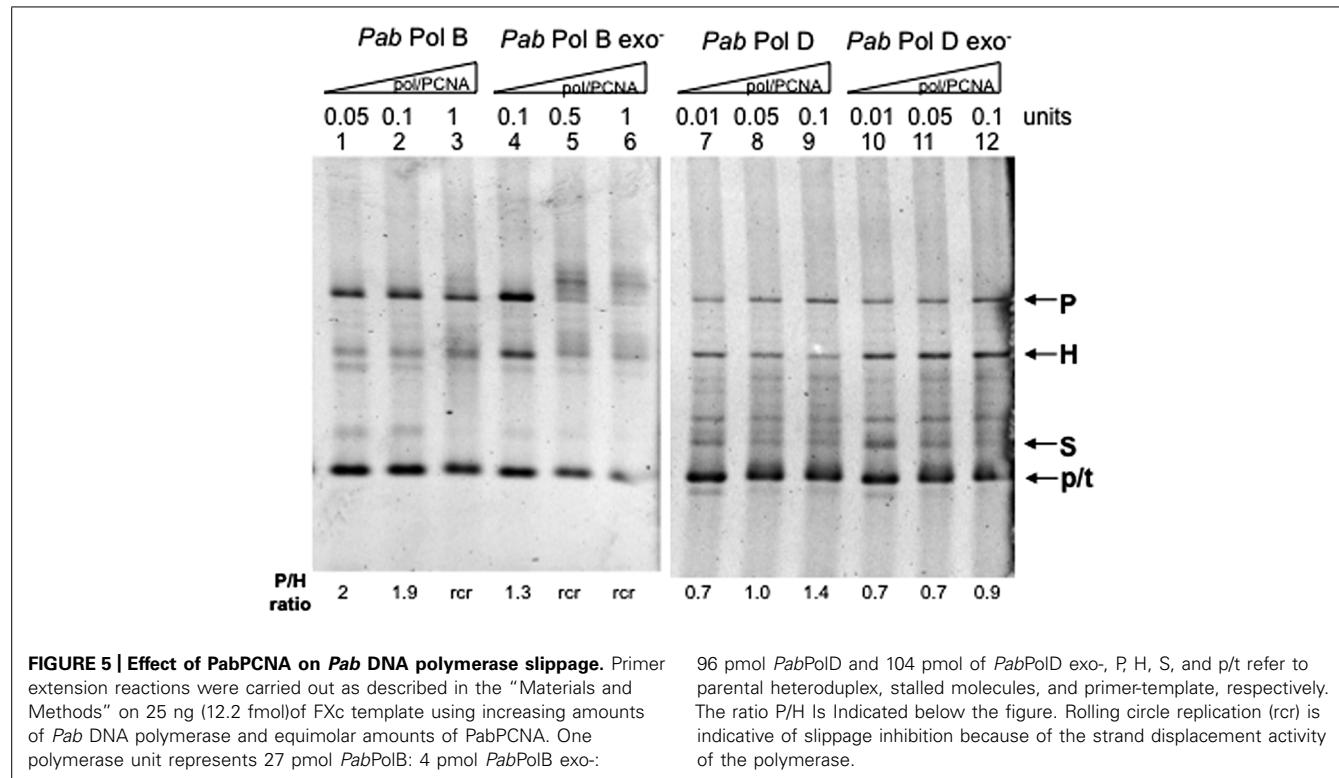
The sliding clamp of Archaea, Eukarya, and Bacteria forms a ring around dsDNA that prevents the dissociation of DNA polymerases from their template, thus enhancing processivity (O'Donnell et al., 2013). Moreover, it acts as a platform that regulates polymerase switching, coupling DNA replication and DNA repair (López de Saro, 2009). The *E. coli* sliding clamp β homodimer subunit

requires the clamp loader (or γ complex) to load it onto the template. The addition of β to primer extension reactions on a hairpin-containing template favors Pol III HE slippage as the synthesis of heteroduplex molecules is stimulated (Canceill and Ehrlich, 1996).

Pyrococcus abyssi, possesses a single processivity clamp, PCNA, that forms a homotrimer (Castreç et al., 2009). In contrast to Bacteria and Eukarya, the archaeal PCNA can be loaded onto DNA without a clamp-loader. *Pab*RF-C and *Pab*PolB, but not *Pab*PolD, enhance PCNA loading. *Pab*RF-C and *Pab*PolB associate with *Pab*PCNA, forming a stable complex on primed DNA (Rouillon et al., 2007).

We therefore investigated the role of PCNA on *Pab*PolB and *Pab*PolD slippage (Figure 5). The addition of equimolar amounts of *Pab*PCNA to the FXc replication assay reduced *Pab*PolB slippage. Formation of parental molecules was stimulated and the proportion of heteroduplex molecules diminished (Figure 5, lanes 1–3) with respect to reactions performed in the absence of PCNA (compare with Figure 3, lanes 1–3). *Pab*PCNA addition also reduced *Pab*PolB exo- slippage (Figure 5, lanes 4–6, compare with Figure 3, lanes 4–6). Furthermore, the addition of PCNA resulted in the appearance of slowly migrating high molecular weight molecules (Figure 5, lanes 5–6). These molecules could be the result of rolling circle replication (*rrc*), which implies that after completion of one round of replication, the newly synthesized strand becomes displaced allowing synthesis to continue (Canceill et al., 1999).

The presence of *Pab*PCNA also increases the proportion of parental versus heteroduplex molecules generated by *Pab*PolD (Figure 5, lanes 7–9), indicating that it also represses slippage by this DNA polymerase. However, *Pab*PCNA had only a slight effect on *Pab*PolD exo- slippage (Figure 5, lanes



10–12), as the proportion of parental molecules was only slightly higher.

We conclude from these experiments that *Pab*PCNA stimulates the ability of *Pab*PolB and *Pab*PolD to replicate through a hairpin structure by inhibiting slippage, with the strongest effect in terms of slippage inhibition observed on *Pab*PolD exo-. These results agree with those obtained by Henneke et al. (2005) in which *Pab*PCNA stimulated the strand displacement activity of *Pab*PolB and *Pab*PolD (Henneke, 2012).

THE *Thermus thermophilus* SINGLE-STRANDED DNA BINDING (SSB) PROTEIN DOES NOT AFFECT SLIPPAGE ERRORS PRODUCED BY *Pab* DNA POLS

The amount of slippage exhibited by different DNA polymerases has been shown to be modulated by SSB proteins (Cancéll and Ehrlich, 1996; Cancéll et al., 1999). *E. coli* SSB stimulates the slippage of Pol III HE, inhibits the slippage of *E. coli* polymerase I and T7 DNA polymerase, and has no effect on *E. coli* pol II or T4 DNA polymerase. On the other hand, T4 SSB protein (gp32) inhibits T4 DNA pol slippage but does not affect the slippage properties of the Pol I Klenow fragment. These contrasting effects of SSB proteins on the same DNA template cannot be understood solely in terms of interaction with DNA, but rather as an interaction between SSB proteins and the different polymerases that alters their strand displacement activity (Cancéll et al., 1999).

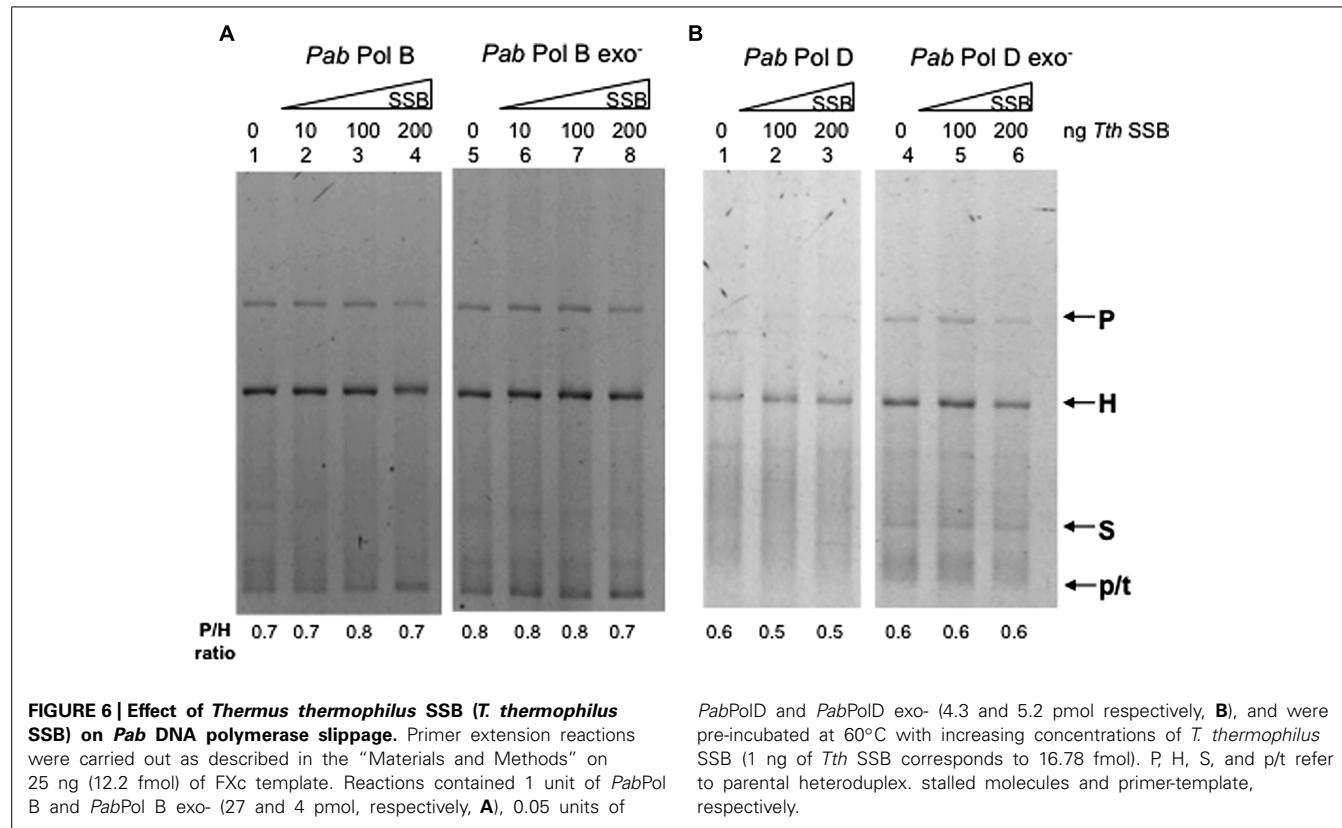
We therefore investigated whether a SSB protein could modify the slippage properties of *Pab* pols. *Thermus thermophilus* (*Tth*) SSB stimulates DNA synthesis of *Tth*DNA polymerase and the heterologous DNA polymerase from the Archaea *P. furiosus* (Perales et al., 2003). Furthermore, *Tth*SSB increases the fidelity of proof-reading deficient *Thermus thermophilus* DNA polymerase (Perales et al., 2003). We tested the slippage properties of *Pab*PolB, *Pab*PolD, and their exo- forms in the presence of increasing amounts of *Tth*SSB. Results obtained for *Pab*PolB and its exo- form are shown in Figure 6A, lanes 1–8. Heteroduplex products were detected in both the presence and absence of *Tth*SSB. Similar results were obtained for *Pab*PolD and its exo- form (Figure 6B, lanes 1–6). We conclude that *Tth*SSB neither stimulated overall synthesis efficiency nor the slippage of *Pab* polymerases under the reaction conditions assayed.

et al., 2003). Furthermore, *Tth*SSB increases the fidelity of proof-reading deficient *Thermus thermophilus* DNA polymerase (Perales et al., 2003). We tested the slippage properties of *Pab*PolB, *Pab*PolD, and their exo- forms in the presence of increasing amounts of *Tth*SSB. Results obtained for *Pab*PolB and its exo- form are shown in Figure 6A, lanes 1–8. Heteroduplex products were detected in both the presence and absence of *Tth*SSB. Similar results were obtained for *Pab*PolD and its exo- form (Figure 6B, lanes 1–6). We conclude that *Tth*SSB neither stimulated overall synthesis efficiency nor the slippage of *Pab* polymerases under the reaction conditions assayed.

DISCUSSION

Interest in DNA repeat instability has increased dramatically since links were established between expansions of trinucleotide repeats and neurodegenerative diseases (Bacolla and Wells, 2009; reviewed in Kim and Mirkin, 2013), microsatellite instability and certain types of cancer (Shah et al., 2010) and the identification of frameshift-mediated regulation of gene expression at simple sequence contingency loci of pathogenic bacteria such as *Neisseria* or *Haemophilus* (Moxon et al., 1994; Bayliss et al., 2001; reviewed in Gemayel et al., 2010).

Because of the association between DNA repeat instability and DNA replication, DNA polymerases have been analyzed *in vitro* to establish their ability to replicate repeated DNA sequences. We have shown previously that the main replicative DNA polymerase of the model bacteria *E. coli*, DNA Pol III HE, is able to slip *in vitro* on hairpin-containing templates despite the high fidelity required for genome replication (Cancéll et al., 1999). Moreover, slippage is stimulated by factors that affect Pol III processivity such as



the presence of β -clamp or SSB proteins (Canceill and Ehrlich, 1996).

Thermostable DNA polymerases are widely used for a number of applications, mostly involving PCR amplification. We have previously shown that replication slippage occurs efficiently even during the first PCR amplification cycle of *Taq* Pol, *Pfu* Pol, PyraTM Pol (*Pab* PolB exo-), or the ExpandTM mixture (*Taq* Pol and *Pwo* Pol; Viguera et al., 2001b). However, no slippage was detected during PCR performed by *Tfu* Pol or Vent[®] Pol. Since, high fidelity DNA polymerases can undergo slippage in terms of base substitution, slippage is only inhibited in those polymerases endowed with high strand displacement activity. Thus, the use of DNA polymerases with high strand displacement activity is advisable when amplifying DNA templates with potential strong secondary structures.

In this study, we have shown that replicative *P. abyssi* DNA polymerases are able to slip *in vitro* on hairpin containing templates as heteroduplex molecules, indicative of slippage error, as well as parental molecules were obtained at every enzyme concentration assayed. This result is quite different to those obtained for the thermostable polymerase B from *P. furiosus* (*Pfu*) or the mesophilic DNA polymerases *E. coli* Pol II and T4, where heteroduplex molecules are the only reaction product (Viguera et al., 2001b). Although native *Pfu* Pol generated some parental molecules, the main reaction products were heteroduplex structures (Figure S2B). In contrast, *Taq* DNA pol, Pol III HE, Pol I, and T7 DNA pol generate heteroduplex and/or parental molecules depending on the polymerase concentration

used in the assay (Canceill et al., 1999; Viguera et al., 2001b; Figure S2A). High polymerase concentrations are believed to promote step-by-step progression inside the hairpin via multiple association/dissociation events (Canceill et al., 1999), thus slippage assay sensitivity to polymerase concentration is consistent with a polymerase possessing some degree of strand displacement activity.

Our interpretation is that the different slippage properties of the closely related *Pab*PolB and *Pfu* Pol are most likely due to *Pab*PolB having higher strand displacement activity, which allows it to generate a higher proportion of parental molecules.

Both *Pab*PolB and *Pab*PolD generated heteroduplex molecules alone or a combination of parental and heteroduplex products at the different magnesium concentration tested whenever synthesis was efficient. This result was somewhat similar to that obtained for *Taq* DNA pol, where either parental or heteroduplex molecules were obtained depending on the magnesium concentration (Viguera et al., 2001b), although we did not find any reaction condition where parental molecules were the sole product of either of the *Pab* polymerases. The strand displacement activity of *Pab*PolB and *Pab*PolD is probably insufficient to reliably open and progress through the hairpin structure even at the lowest magnesium concentration tested; a condition that should reduce DNA duplex stability. Consequently, even if *Pab*PolB and *Pab*PolD have different roles at the replication fork, they do not differ in terms of their slippage properties. This result prompted us to study other cellular factors that could inhibit the slippage errors detected by our *in vitro* assay.

We have shown that the *PabPCNA* sliding clamp promotes the synthesis of parental molecules by *PabPolB*. This effect is even more prominent for the exonuclease-deficient *PabPolB* (**Figure 5**). In comparison to *PabPolB*, inhibition of slippage by PCNA was weaker for *PabPolD* (**Figure 5**). *PabPolB* has been identified as the leading strand DNA polymerase (Henneke et al., 2005; Rouillon et al., 2007). *PabPCNA* interacts with *PabPolB* in a DNA-dependent way and stimulates its processivity, clamping *PabPolB* to DNA (Henneke et al., 2005; Henneke, 2012). The higher processivity of the *PabPCNA-PolB* complex and stimulation of strand displacement activity (Henneke et al., 2005) would facilitate opening of the DNA duplex leading to the synthesis of parental molecules. The effect of *PabPCNA* is further increased for *PabPolB* exo-. One possibility is that strand displacement activity is increased to some degree in this mutant and that this facilitates parental formation as has been observed for T7 pol exo- (SequenaseTM) and Pol II exo- (Canceill et al., 1999).

In our opinion, this result confirms *PabPCNA-PolB* as a competent and stable complex, capable of continuously synthesizing the leading strand. Upon encountering secondary structures (such as hairpin loops), DNA synthesis is unperturbed and the *PabPCNA-PolB* complex is capable of continuing strand elongation.

The reason that *PabPCNA* inhibited *PabPolD* replication slippage to a lesser extent than *PabPolB* is probably due to insufficient stimulation of strand displacement activity (Henneke et al., 2005) under the conditions tested. *PabPCNA* binds PolB and PolD in different ways (Castrec et al., 2009). Two PCNA-interacting protein (PIP) boxes are needed for *PabPolD* binding to *PabPCNA* whereas only one PIP motif is essential for *PabPolB* binding. This suggests that the mechanism involved in the *PabPCNA*-mediated stimulation of *PabPolD* may be different from that involved in *PabPolB*.

Previous work (Canceill and Ehrlich, 1996) has shown that in *E. coli*, the β-clamp does not stimulate the generation of parental molecules by Pol III HE but instead increases the formation of heteroduplex ones. Our finding that the addition of *PabPCNA* promotes parental formation by *PabPolB* and *PabPolD*, indicates that interaction between DNA polymerases and PCNA in the archaeal *P. abyssi* promotes faithful replication of DNA secondary structures. The functional homology between archaeal and eukaryal proteins, i.e., human PCNA can be loaded onto DNA by the *P. abyssi* RF-C complex (Henneke et al., 2002), suggests that slippage by replicative eukaryal DNA polymerases may also be inhibited by presence of the sliding clamp.

ACKNOWLEDGMENTS

We thank Dr. C. Perales and Dr. J. Berenguer (CBM-SO, Madrid) for their generous gift of *TthSSB* and Dr. J. R. Pearson for scientific English editing of the manuscript. This work was supported by grants BFU2007-64153 from the Ministerio de Educación y Ciencia and P09-CVI-5428 from the Junta de Andalucía to Enrique Viguera. Ghislaine Henneke was financially supported by grant ANR-10-JCJC-1501 01 from the National Research Agency. Melissa Castillo-Lizardo acknowledges the short-term fellowship from EMBO and the FPI predoctoral Fellowship BES-2005-10150 from Ministerio de Educación y Ciencia, Spain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00403/abstract>

REFERENCES

- Aguilera, A., and Gomez-Gonzalez, B. (2008). Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* 9, 204–217. doi: 10.1038/nrg2268
- Bacolla, A., and Wells, R. D. (2009). Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.* 48, 273–285. doi: 10.1002/mc.20507
- Barry, E. R., and Bell, S. D. (2006). DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.* 70, 876–887. doi: 10.1128/MMBR.00029-06
- Bayliss, C. D., Field, D., and Moxon, E. R. (2001). The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J. Clin. Invest.* 107, 657–662. doi: 10.1172/JCI12557
- Bierne, H., Ehrlich, S. D., and Michel, B. (1991). The replication termination signal terB of the *Escherichia coli* chromosome is a deletion hot spot. *EMBO J.* 10, 2699–2705.
- Bierne, H., Ehrlich, S. D., and Michel, B. (1996). Deletions at stalled replication forks occur by two different pathways. *EMBO J.* 16, 3332–3340. doi: 10.1093/emboj/16.11.3332
- Canceill, D., and Ehrlich, S. D. (1996). Copy-choice recombination mediated by DNA polymerase III holoenzyme from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6647–6652. doi: 10.1073/pnas.93.13.6647
- Canceill, D., Viguera, E., and Ehrlich, S. D. (1999). Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency. *J. Biol. Chem.* 274, 27481–27490. doi: 10.1074/jbc.274.39.27481
- Castrec, B., Rouillon, C., Henneke, G., Flament, D., Querellou, J., and Raffin, J. P. (2009). Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. *J. Mol. Biol.* 394, 209–218. doi: 10.1016/j.jmb.2009.09.044
- Cline, J., Braman, J., and Hogrefe, H. (1996). PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucl. Acids Res.* 24, 3546–3551. doi: 10.1093/nar/24.18.3546
- d'Alencon, E., Petranovic, M., Michel, B., Noirot, P., Aucourturier, A., Uzest, M., et al. (1994). Copy-choice illegitimate DNA recombination revisited. *EMBO J.* 13, 2725–2734.
- Eckert, K. A., and Kunkel, T. A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucl. Acids Res.* 18, 3739–3744. doi: 10.1093/nar/18.13.3739
- Engler, M. J., Lechner, R. L., and Richardson, C. C. (1983). Two forms of the DNA polymerase of bacteriophage T7. *J. Biol. Chem.* 258, 11165–11173.
- Feschenko, V. V., and Lovett, S. T. (1998). Slipped misalignment mechanisms of deletion formation: analysis of deletion endpoints. *J. Mol. Biol.* 276, 559–569. doi: 10.1006/jmbi.1997.1566
- Gemayel, R., Vinces, M. D., Legendre, M., and Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* 44, 445–477. doi: 10.1146/annurev-genet-072610-155046
- Greenstein, D., and Horiuchi, K. (1987). Interaction between the replication origin and the initiator protein of the filamentous phage f1. Binding occurs in two steps. *J. Mol. Biol.* 197, 157–174. doi: 10.1016/0022-2836(87)90115-X
- Gueguen, Y., Rolland, J. L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., et al. (2001). Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur. J. Biochem.* 268, 5961–5969. doi: 10.1046/j.0014-2956.2001.02550.x
- Henneke, G. (2012). In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways. *Biochem. J.* 447, 271–280. doi: 10.1042/BJ20120959
- Henneke, G., Flament, D., Hubscher, U., Querellou, J., and Raffin, J. P. (2005). The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J. Mol. Biol.* 350, 53–64. doi: 10.1016/j.jmb.2005.04.042
- Henneke, G., Gueguen, Y., Flament, D., Azam, P., Querellou, J., Dietrich, J., et al. (2002). Replication factor C from the hyperthermophilic archaeon *Pyrococcus*

- abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RFC PCNA-binding domain. *J. Mol. Biol.* 323, 795–810. doi: 10.1016/S0022-2836(02)01028-8
- Ishino, Y., Komori, K., Cann, I. K., and Koga, Y. (1998). A novel DNA polymerase family found in Archaea. *J. Bacteriol.* 180, 2232–2236.
- Kim, J. C., and Mirkin, S. M. (2013). The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* 23, 280–288. doi: 10.1016/j.gde.2013.04.009
- Kroutil, L. C., Register, K., Bebenek, K., and Kunkel, T. A. (1996). Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry* 35, 1046–1053. doi: 10.1021/bi952178h
- Kunkel, T. A., and Bebenek, K. (2000). DNA replication fidelity. *Annu. Rev. Biochem.* 69, 497–529. doi: 10.1146/annurev.biochem.69.1.497
- Lechner, R. L., Engler, M. J., and Richardson, C. C. (1983). Characterization of strand displacement synthesis catalyzed by bacteriophage T7 DNA polymerase. *J. Biol. Chem.* 258, 11174–11184.
- López de Saro, F. J. (2009). Regulation of interactions with sliding clamps during DNA replication and repair. *Curr. Genomics* 10, 206–215. doi: 10.2174/138920209788185234
- Lovett, S. T. (2004). Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52, 1243–1253. doi: 10.1111/j.1365-2958.2004.04076.x
- Lyons-Darden, T., and Topal, M. (1999). Effects of temperature, Mg²⁺ concentration and mismatches on triplet-repeat expansion during DNA replication in vitro. *Nucl. Acids Res.* 27, 2235–2240. doi: 10.1093/nar/27.11.2235
- Madsen, C. S., Ghivizzani, S. C., and Hauswirth, W. W. (1993). In vivo and in vitro evidence for slipped mispairing in mammalian mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 90, 7671–7675. doi: 10.1073/pnas.90.16.7671
- McMurray, C. T. (1999). DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1823–1825. doi: 10.1073/pnas.96.5.1823
- Michel, B. (2000). Replication fork arrest and DNA recombination. *Trends Biochem. Sci.* 25, 173–178. doi: 10.1016/S0968-0004(00)01560-7
- Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature* 447, 932–940. doi: 10.1038/nature05977
- Moxon, E. R., Rainey, P. B., Nowak, M. A., and Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* 4, 24–33. doi: 10.1016/S0960-9822(00)00005-1
- O'Donnell, M., Langston, L., and Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.* 5, pii: a010108. doi: 10.1101/cshperspect.a010108
- Palud, A., Villani, G., L'Haridon, S., Querellou, J., Raffin, J. P., and Henneke, G. (2008). Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance? *Mol. Microbiol.* 70, 746–761. doi: 10.1111/j.1365-2958.2008.06446.x
- Pearson, C. E., Nichol Edamura, K., and Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729–742. doi: 10.1038/nrg1689
- Perales, C., Cava, F., Meijer, W. J., and Berenguer, J. (2003). Enhancement of DNA, cDNA synthesis and fidelity at high temperatures by a dimeric single-stranded DNA-binding protein. *Nucl. Acids Res.* 31, 6473–6480. doi: 10.1093/nar/gkg865
- Raymann, K., Forterre, P., Brochier-Armanet, C., and Gribaldo, S. (2014). Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in Archaea. *Genome Biol. Evol.* 6, 192–212. doi: 10.1093/gbe/evu004
- Richards, R. I., and Sutherland, G. R. (1992). Dynamic mutations: a new class of mutations causing human disease. *Cell* 70, 709–712. doi: 10.1016/0092-8674(92)90302-S
- Rouillon, C., Henneke, G., Flament, D., Querellou, J., and Raffin, J. P. (2007). DNA Polymerase Switching on Homotrimeric PCNA at the Replication Fork of the Euryarchaeon *Pyrococcus abyssi*. *J. Mol. Biol.* 369, 343–355. doi: 10.1016/j.jmb.2007.03.054
- Samadashwily, G. M., Gordana, R., and Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication in vivo. *Nat. Genet.* 17, 298–304. doi: 10.1038/ng1197-298
- Shah, S. N., Hile, S. E., and Eckert, K. A. (2010). Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes. *Cancer Res.* 70, 431–435. doi: 10.1158/0008-5472.CAN-09-3049
- Sinden, R. R., Pytlos-Sinden, M. J., and Potaman, V. N. (2007). Slipped strand DNA structures. *Front. Biosci.* 12:4788–4799. doi: 10.2741/2427
- Soengas, M. S., Esteban, J. A., Lazaro, J. M., Bernard, A., Blasco, M. A., Salas, M., et al. (1992). Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. *EMBO J.* 11, 4227–4237.
- Tran, H. T., Degtyareva, N. P., Koloteva, N. N., Sugino, A., Masumoto, H., Gordekin, D. A., et al. (1995). Replication slippage between distant short repeats in *Saccharomyces cerevisiae* depends on the direction of replication and the RAD50 and RAD52 genes. *Mol. Cell. Biol.* 15, 5607–5617.
- Trinh, T. Q., and Sinden, R. R. (1993). The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* 134, 409–422.
- Viguera, E., Canceill, D., and Ehrlich, S. D. (2001a). Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J.* 20, 2587–2595. doi: 10.1093/emboj/20.10.2587
- Viguera, E., Canceill, D., and Ehrlich, S. D. (2001b). In vitro replication slippage by DNA polymerases from thermophilic organisms. *J. Mol. Biol.* 312, 323–333. doi: 10.1006/jmbi.2001.4943

Conflict of Interest Statement: The Review Editor Bernard Connolly declares that, despite having collaborated and published with Ghislaine Henneke, the review process was handled objectively. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 May 2014; accepted: 17 July 2014; published online: 07 August 2014.

Citation: Castillo-Lizardo M, Henneke G and Viguera E (2014) Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota *Pyrococcus abyssi*. *Front. Microbiol.* 5:403. doi: 10.3389/fmicb.2014.00403

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Castillo-Lizardo, Henneke and Viguera. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



PCR performance of a thermostable heterodimeric archaeal DNA polymerase

Tom Killelea^{1,2,3}, Céline Ralec^{1,2,3}, Audrey Bossé^{1,2,3} and Ghislaine Henneke^{1,2,3*}

¹ Université de Bretagne Occidentale, UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes, Plouzané, France

² Ifremer, UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes, Plouzané, France

³ CNRS, UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes, Plouzané, France

Edited by:

Zvi Kelman, University of Maryland,
USA

Reviewed by:

Uli Stigl, King Abdullah University
of Science and Technology, Saudi
Arabia

Dennis W. Grogan, University of
Cincinnati, USA

***Correspondence:**

Ghislaine Henneke, Laboratoire de
Microbiologie des Environnements
Extrêmes, Ifremer, UMR 6197, Zi de
la point du diable, CS 10070, 29280
Plouzané, France
e-mail: ghenneke@ifremer.fr

DNA polymerases are versatile tools used in numerous important molecular biological core technologies like the ubiquitous polymerase chain reaction (PCR), cDNA cloning, genome sequencing, and nucleic acid based diagnostics. Taking into account the multiple DNA amplification techniques in use, different DNA polymerases must be optimized for each type of application. One of the current tendencies is to reengineer or to discover new DNA polymerases with increased performance and broadened substrate spectra. At present, there is a great demand for such enzymes in applications, e.g., forensics or paleogenomics. Current major limitations hinge on the inability of conventional PCR enzymes, such as *Taq*, to amplify degraded or low amounts of template DNA. Besides, a wide range of PCR inhibitors can also impede reactions of nucleic acid amplification. Here we looked at the PCR performances of the proof-reading D-type DNA polymerase from *P. abyssi*, Pab-polD. Fragments, 3 kilobases in length, were specifically PCR-amplified in its optimized reaction buffer. Pab-polD showed not only a greater resistance to high denaturation temperatures than *Taq* during cycling, but also a superior tolerance to the presence of potential inhibitors. Proficient proof-reading Pab-polD enzyme could also extend a primer containing up to two mismatches at the 3' primer termini. Overall, we found valuable biochemical properties in Pab-polD compared to the conventional *Taq*, which makes the enzyme ideally suited for cutting-edge PCR-applications.

Keywords: DNA polymerase, Archaea, family D, PCR, *Pyrococcus*

INTRODUCTION

On the basis of their amino acid sequence and structural analysis, DNA polymerases have been classified into seven families, A, B, C, D, E, X, and Y (Delarue et al., 1990; Braithwaite and Ito, 1993; Joyce and Steitz, 1994; Cann et al., 1998; Ishino et al., 1998; Ohmori et al., 2001; Lipps et al., 2003). Despite the conserved template-directed synthesis (or editing) of a complementary deoxyribonucleotide chain (Kornberg and Baker, 1992; Hubscher et al., 2010) and the similarity of the three-dimensional organization of their polymerase domain ("palm," "thumb," and "finger") (Joyce and Steitz, 1994; Rothwell and Waksman, 2005), DNA polymerases differ extensively in many of their specific features (e.g., processivity, fidelity, rate of DNA synthesis, and nucleotide selectivity) (Hubscher et al., 2010; Langhorst et al., 2012).

Beginning with the discovery and characterization of DNA polymerase I (Family A) from *Thermus aquaticus* (*Taq*) (Chien et al., 1976), a variety of thermostable DNA polymerases have been isolated and identified from prokaryotic organisms. Besides their crucial biological functions, thermostable DNA polymerases have proven to be technically and economically important enzymes. They are versatile tools used in DNA technologies such as cycle sequencing and polymerase chain reaction (PCR) (Pavlov et al., 2004). Since its invention by Saiki et al. (1985), PCR has become a widespread molecular biology method. Originally

PCR was developed to specifically amplify a stretch of DNA prior to cloning; however, its flexibility underpins a number of applications such as site-directed mutagenesis, genetic diagnostics, gene therapy, forensics, and paleogenomics.

In the PCR, DNA amplification is performed by thermostable enzymes; invariably either family A DNA polymerases from thermophilic and hyperthermophilic Bacteria (e.g., *Thermus aquaticus*, *Taq*-polA and *Thermotoga maritima*, *Tma*-polA) or family B DNA polymerases from hyperthermophilic Archaea (e.g., *Pyrococcus furiosus*, *Pfu*-polB and *Pyrococcus abyssi*, *Pab*-polB; *Isis*™). Family Y DNA polymerase from the hyperthermophilic archaeon *Sulfolobus solfataricus*, *Sso*-polY, is also an enzyme marketed for PCR, but with specialist applications (McDonald et al., 2006). Each thermostable DNA polymerases has different characteristics (e.g., thermostability, processivity, fidelity, specificity, modified nucleotides selection, resistance to contaminants and inhibitors, slippage, pyrophosphorylation) and to achieve optimal results, the choice of a PCR enzyme depends on the application itself (e.g., high-yield PCR, high-fidelity PCR, routine PCR, multiplex PCR, colony PCR, difficult PCR, long Range PCR, fast PCR, incorporation of modified nucleotides). Detailed information about individual properties of PCR enzymes and their related applications have been recently reviewed (Terpe, 2013).

Archaeal family B DNA polymerases are generally more thermostable enzymes than bacterial polymerases (Kong et al., 1993; Takagi et al., 1997; Cambon-Bonavita et al., 2000; Gueguen et al., 2001; Hogrefe et al., 2001; Moussard et al., 2006; Marsic et al., 2008). In addition, when accuracy is a desired property it leads to a preference for archaeal enzymes which possess a 3'-5' proof-reading exonuclease activity, absent in most thermostable bacterial polymerases (Eckert and Kunkel, 1991; Cline et al., 1996; Perler et al., 1996). On the other hand, archaeal family B DNA polymerases can incorporate dUTP during DNA replication but cannot copy these strands in subsequent DNA amplification rounds (Fogg et al., 2002). With constant development of new techniques based on PCR, improved DNA polymerase variants are continuously being engineered including: polymerases which are thermo-activated (hot start polymerases) (Sharkey et al., 1994), bacterial and archaeal DNA polymerase derivatives with increased processivity (Wang et al., 2004), archaeal family B DNA polymerase variants insensitive to uracil inhibition (Fogg et al., 2002), or thermostable enzymes proficient in the synthesis of fluorescent CyDNAs (Ghadessy et al., 2004; Wynne et al., 2013). Error-prone PCR which creates random mutagenesis of the parental gene relies on different strategies using either low-fidelity DNA polymerase variants (Biles and Connolly, 2004) or error-prone PCR conditions (McCullum et al., 2010; Le et al., 2013). Most of these conventional marketed PCR enzymes have the drawback of replicating exclusively native DNAs. Among thermostable DNA polymerases that can counteract this major limitation, archaeal translesional family Y DNA polymerases are of particular interest. They are able to bypass a variety of DNA lesions and therefore are well-suited for the PCR amplification of ancient and damaged DNAs (McDonald et al., 2006).

Over 14 years ago a new family of archaeal DNA polymerases (the D-family) was discovered (Uemori et al., 1997). Despite obvious interest in the biochemical characteristics of archaeal family D DNA polymerases (Cann et al., 1998; Gueguen et al., 2001) there is limited information regarding the structure (Yamasaki et al., 2010; Matsui et al., 2011) and kinetics of these enzymes (Palud et al., 2008; Richardson et al., 2013a). However, it is known that these thermostable DNA polymerases are heterodimeric and comprise a small subunit (DP1), possessing 3' → 5' exonuclease activity, and a large subunit (DP2), exhibiting DNA polymerase activity (Cann and Ishino, 1999). The small subunit shares low level of homology with the non-catalytic B-subunits of the eukaryotic family B DNA polymerases (Cann et al., 1998; Ishino et al., 1998; Gueguen et al., 2001). In contrast, the sequence of the large subunit shows no significant homology to any other DNA polymerase (Macneill et al., 2001). Currently, the growing body of evidence suggests involvement of the family D DNA polymerases in genome replication in Archaea (Henneke et al., 2005; Rouillon et al., 2007; Castrec et al., 2009; Cubanova et al., 2013).

Family D DNA polymerases from hyperthermophilic Archaea which have been biochemically characterized, to date, are from the *Pyrococcus* genus such as *Pyrococcus horikoshii* (Shen et al., 2001), *Pyrococcus furiosus* (Uemori et al., 1997), and *Pyrococcus abyssi* (Gueguen et al., 2001). These microorganisms contain only one family B enzyme in addition to the family D DNA polymerase. In contrast with commercialized family B enzymes

(Pfu-polB and Pab-polB), none of the family D DNA polymerases have been reported as active enzymes in PCR or in other DNA technologies.

Family D DNA polymerase from *Pyrococcus abyssi* shows comparable nucleotide selectivity to family B, and increased fidelity with the active proofreading (Palud et al., 2008; Richardson et al., 2013a). Family D DNA polymerase preferentially binds to primer/template with an affinity higher than family B, while showing reduced DNA synthesis of smaller DNA fragments (Henneke et al., 2005). The assembly of the two subunits into a heterodimer is required to substantially increase both polymerase and exonuclease activities in family D, while both activities are contained within the same polypeptide in the family B DNA polymerase (Castrec et al., 2010; Gouge et al., 2012). These functional properties suggest that family D DNA polymerase might perform PCR performance distinct than Pab-polB. In this paper, the ability of the recombinant family D DNA polymerase from *Pyrococcus abyssi* (Pab-polD) to PCR-amplify DNA has been developed in terms of biochemical and PCR performance parameters (e.g., stability to heat denaturation steps, extension efficiency, resistance to common PCR inhibitors). These results are compared with data acquired from commercial thermostable DNA polymerases (Pab-polB and Taq-polA) and reveal that family D DNA polymerase has significant commercial value in PCR technology.

MATERIALS AND METHODS

CHEMICALS AND ENZYMES

Unlabeled dNTPs were purchased from MP Biomedicals. Pab-polD was cloned, expressed, and purified as described (Henneke et al., 2005). One unit of Pab-polD corresponds to the incorporation of 1 nmol of total dTMP into acid precipitable material per minute at 65°C in a standard assay containing 0.5 mg (nucleotides) of poly(dA)/oligo(dT)_{10:1}. Pab-polB (*Isis* DNA polymerase) and Taq-polA (*Taq* DNA polymerase) were purchased from MP Biomedicals. All other chemicals and bioreagents were analytical grade and purchased from Sigma-Aldrich (St. Louis, MO). Bioactive small molecules (Human hemoglobin, humic acid, hematin, heparin, and urea) were molecular biology grade from Sigma-Aldrich (St. Louis, MO). The 1.7 million base-pair genome of *Pyrococcus abyssi* GE5 was obtained as described (Charbonnier et al., 1995).

POLYMERASE CHAIN REACTION (PCR ENZYMES)

PCR primers for the amplification of targets in genomic DNA from *P. abyssi* (Pab) genomic sequence from 1323272 to 1333272 base pairs (bp) were purchased from Eurogentec (Belgium). The primer sequences, the Pab genomic sequence, and the size of the expected amplicons (in kilobases, kb) are summarized in **Table 1**. These selective amplifications were dictated by the availability of total genomic DNA from *P. abyssi* devoid of any potential PCR inhibitors and the use of thermally stable oligonucleotide primers. PCR performance parameters of Pab-polD were determined in the optimized buffer: 20 mM Tris-HCl pH 9, 25 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgCl₂, 0.1 mg/ml Bovine Serum Albumin (BSA), 0.1% (v/v) Tween 20. PCR reactions (25 µl) contained 200 nM of each primer, 200 µM dNTPs, and 100 ng

Table 1 | Primers applied in this study.

Primer name	Target length (kb)	Primer sequence (5'-3')	Pab genomic sequence (bp)	GC % of DNA target
Start_F		GCAAAGGAGTTGCCAGCTTATAGAGACGGCC	1333240–1333272	
Start_1M_F		GCAAAGGAGTTGCCAGCTTATAGAGACGGCA	1333240–1333272	
Start_2M_F		GCAAAGGAGTTGCCAGCTTATAGAGACGGAA	1333240–1333272	
Start_3M_F		GCAAAGGAGTTGCCAGCTTATAGAGACGCAA	1333240–1333272	
Start_4M_F		GCAAAGGAGTTGCCAGCTTATAGAGACCCAA	1333240–1333272	
0.5_R	0.5	TCAACCTCCTGGGTTCACCTCGGCCCTC	1332773–1332802	47.2
1_R	1.1	CCTCAGCCTAACGCTTATTGGTTCCCCGT	1332173–1332202	47.1
2_R	1.95	GAAGATTAAGGGAGAGGGGAGGGCAAAGGCTGGTGG	1331322–1331357	46.8
3_R	2.95	CGAGGAAGCTAGAGGAGAGGGTGCTGAGTGC	1330322–1330353	47.5
4_R	4.15	GGAGGGGCAACTATGGATACCACCCCTTC	1329122–1329153	46.7
10_R	10	TCCAAGAGTTCCCTCCCTGGGAAACGGCTGAACC	1323273–1323306	47

F and R denote forward and reverse primers, respectively. M stands for mismatch.

of genomic DNA unless otherwise specified. The PCR conditions for commercial Taq-polA and Pab-polB were set according to the manufacturers' instructions. All reactions were run in (at least) duplicate. Negative control included all reaction components without genomic DNA. The amplification was carried out in GeneAmp® PCR System 9700 Thermal Cycler (Applied Biosystems) and in Veriti® 96-Well Thermal Cycler (Applied Biosystems). Cycling conditions were 2 min at 94°C; 30 cycles with 1 min denaturation at 94°C, 1 min annealing at 58°C and extension at 72°C at the indicated times. A final extension step at 72°C was applied before the termination of the reaction as specified in the corresponding figure legends. Elongation temperature was set at 72°C according to the manufacturer protocols for Taq-PolA and Pab-polB, therefore validating the temperature of assay performance by Pab-polD. The products were analyzed with 1% agarose gel electrophoresis, stained with ethidium bromide, and visualized with the Molecular Imager FX (BioRad). When mentioned, activity (%) is expressed as a percentage of the maximal value obtained in each experiment.

PCR experiments in the presence of inhibitors were conducted with the optimized Pab-polD buffer as described above. The PCR conditions for Pab-polB and Taq-polA were set according to the manufacturers' instructions. A 0.5 kb fragment was amplified from *Pab* genomic DNA using the 500 bp reverse and forward primers (listed in **Table 1**). Titration of each inhibitor was performed at least in triplicate. Cycling conditions were 2 min at 94°C; 30 cycles with 1 min denaturation at 94°C, 1 min annealing at 58°C, and 2 min extension at 72°C; final extension, 5 min at 72°C. The products were analyzed with 1% agarose gel electrophoresis, stained with ethidium bromide, and visualized with the Molecular Imager FX (BioRad).

RESULTS

OPTIMIZED PCR REACTION CONDITIONS

The optimized buffer for PCR with Pab-polD was obtained by varying different components of the standard Pab-polB reaction buffer. Pab-polD PCR activity was optimal in 10–30 mM Tris-HCl buffer concentration (**Figure 1A**) and between pH 8.3 and 9 (measured at 25°C) (**Figure 1B**). Incubation of Pab-polD

with either magnesium chloride (MgCl₂) or magnesium sulfate (MgSO₄) in the same concentration range resulted in the amplification of non-specific and undesirable PCR products with MgSO₄ (**Figure 1C**). Reactions carried out with MgCl₂ gave rise to the amplification of specific products in the optimal concentration range tested (**Figure 1D**). The effects of different salt concentrations of potassium chloride (KCl) (**Figure 1E**) and ammonium sulfate ((NH₄)₂SO₄) (**Figure 1F**) were analyzed in PCR by Pab-polD. The maximal activity detected with the two salts was 0–20 mM for KCl and 15–25 mM for (NH₄)₂SO₄, respectively. Although the presence of (NH₄)₂SO₄ further enhanced PCR amplification at optimal concentration, KCl could be dispensable. Finally, the two additives, Tween and BSA, used in the standard Pab-polB reaction buffer were added or omitted in Pab-polD PCR reactions. In the conditions tested, 0.1 mg/ml BSA and 0.1% Tween did not significantly improve the amount of PCR products by Pab-polD (**Figure 1G**). Overall, the optimal reaction buffer for *in vitro* amplification of DNA fragments by Pab-polD has been determined and is now available in **Table 2**.

EFFECT OF INPUT GENOMIC DNA ON PCR EFFICIENCY AND SPECIFICITY

PCR amplification, targeting the 0.5 kb fragment in the 1.7 million base-pair genome of *P. abyssi* (**Table 1**), was employed to determine the minimal amount of DNA required. In its optimal reaction conditions, Pab-polD was able to specifically amplify the 0.5 kb target from 0.5 to 100 ng of input genomic DNA (**Figure 2A**). Although the yield of PCR products was severely reduced at 0.5–1 ng, all three enzymes retained polymerase activity (~2–5% of activity) (**Figures 2A–C**). In the presence of 0.1 ng, only Pab-polB was capable of amplification of the 0.5 kb DNA target (**Figure 2B**).

IMPACT OF THERMAL DENATURATION DURING CYCLING

The resistance of Pab-polD to the temperature of the denaturation step during cycling was investigated in comparison with Taq-polA and Pab-polB. PCR amplifications of the 0.5 kb DNA target were performed with 4 different thermal denaturation steps during cycling (91, 95, 97, and 99°C). As shown in **Figure 3A**,

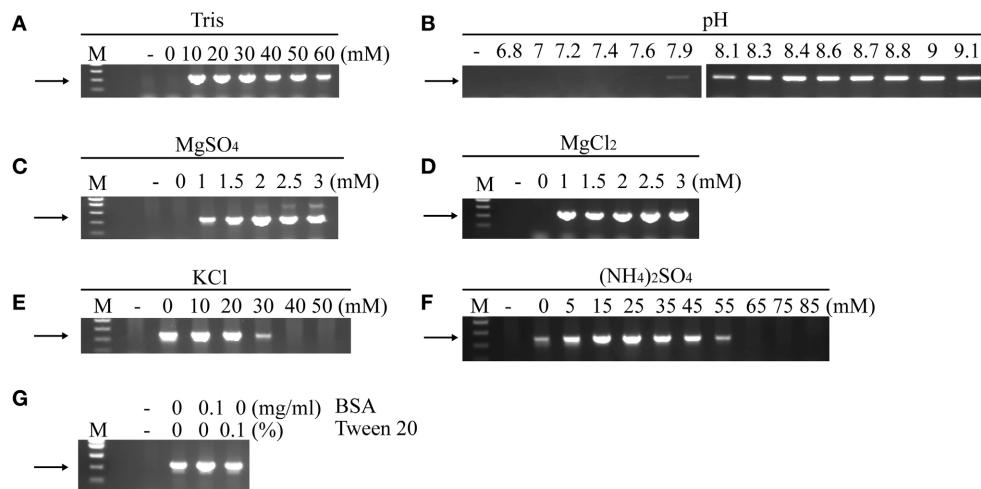


FIGURE 1 | Effect of buffer concentration, pH, salts, ions, BSA, and Tween 20 on PCR amplification of the 0.5 kb target by Pab-poID. PCR reactions were carried out using 100 ng of genomic DNA with 0.1 U of Pab-poID. The optimized reaction buffer was used as is shown in **Table 2**, except for the variable condition, which was altered as shown in the figure. PCR program was (2 min at 94°C) × 1; (1 min at 94°C,

1 min at 58°C, 2 min at 72°C) × 30; (5 min at 72°C) × 1. Effect of varying the Tris concentration (**A**), pH (**B**), MgSO₄ (**C**), MgCl₂ (**D**), KCl (**E**), (NH₄)₂SO₄ (**F**). The absence or presence of BSA and Tween 20 are shown in (**G**). Molecular weight markers (M) are SmartLadder SF from Eurogentec. (–) denotes negative control without Pab-poID. The arrow indicates the specific 0.5 kb band.

Table 2 | Properties of experimental thermophilic DNA polymerases.

Pab-poID	Pab-polB	Taq-polA
Name	Pab-poID	<i>Isis</i>
Microorganism	<i>Pyrococcus abyssi</i>	<i>Thermus aquaticus</i>
Pol family	D	B
Half-life	90°C/50 min (Gueguen et al., 2001) 80°C/2 h (Gueguen et al., 2001)	100°C/5 h 90°C/10 h (Gueguen et al., 2001; Dietrich et al., 2002)
Error rate × 10 ⁻⁶ (/bp/duplication)	N.P.	0.66 (Dietrich et al., 2002)
Reaction buffer	20 mM Tris-HCl, pH (25°C) 9.0, 25 mM KCl, 10 mM (NH ₄) ₂ SO ₄ , 2 mM MgCl ₂ , 0.1% Tween 20, 0.1 mg/ml BSA	20 mM Tris-HCl, pH (25°C) 9.0, 25 mM KCl, 10 mM (NH ₄) ₂ SO ₄ , 1.5 mM MgSO ₄ , 0.1% Tween 20, 0.1 mg/ml BSA
Exonuclease activities	3'-5' (Gueguen et al., 2001)	3'-5' (Gueguen et al., 2001)
Extension rate: kb/min (nt/s)	0.33 (5.5)	0.48 (8)
24 (Dietrich et al., 2002)		
10 mM Tris-HCl pH (25°C) 9.0, 50 mM KCl, 1.5 mM, MgCl ₂ , 0.2 mg/mL BSA		
5'-3' (Lawyer et al., 1993)		

N.P., Not Published.

Error rate for Pab-polB and Taq-polA were determined in the same experiments under the same conditions (Dietrich et al., 2002).

Pab-poID could yield specific PCR products (15% activity compared to 91°C) when the denaturation temperature was as high as 97°C. PCR products were hardly detectable at 99°C. Interestingly, the PCR efficiency of Pab-polB was not profoundly affected by the increase in the thermal denaturation step during cycling, while Taq-polA was mostly inactive above 95°C (**Figures 3B,C**). Taken together, these results are in agreement with those published previously (Dietrich et al., 2002) and indicate that hyperthermophilic Pab DNA polymerases are more robust than the thermophilic Taq-polA.

RATE OF DNA EXTENSION

In order to determine the rate of primer extension by Pab-poID, the 1.95 kb DNA target was amplified using variants of the endpoint PCR method employed throughout this study. Numerous reactions were carried out, with each PCR possessing an incrementally larger extension time than the last, until end-point PCR products were detectable on an agarose gel. Our initial attempt, in which the longest extension time was set to 240 s, failed with Pab-poID and Taq-polA (**Figures 4A,C**). However, the 1.95 kb DNA target was successfully amplified

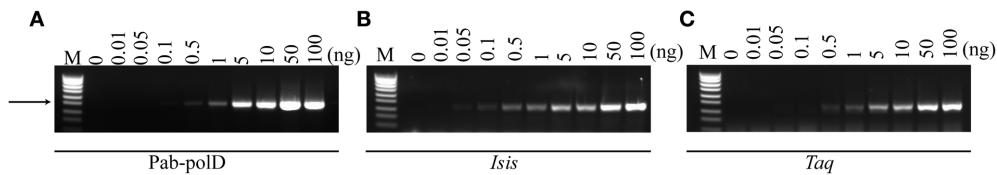


FIGURE 2 | Effect of input genomic DNA on PCR efficiency and specificity. PCR amplification of the 0.5 kb target (**Table 1**) was carried out with 0.1 U of Pab-polD (**A**), 1 U of Pab-polB (**B**), and 1 U of Taq-polA (**C**) in their respective reaction buffer (**Table 2**). PCR program

was (2 min at 94°C) × 1; (1 min at 94°C, 1 min at 58°C, 2 min at 72°C) × 30; (5 min at 72°C) × 1. Molecular weight markers (M) are SmartLadder SF from Eurogentec. The arrow indicates the specific 0.5 kb band.

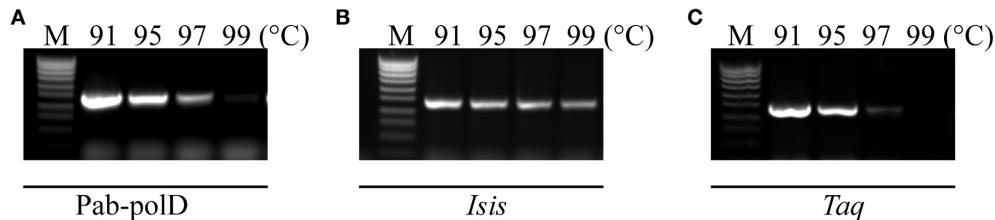


FIGURE 3 | Impact of thermal denaturation during cycling. PCR amplification of the 0.5 kb target (**Table 1**) was carried out with 0.1 U of Pab-polD (**A**), 1 U of Pab-polB (**B**), and 1 U of Taq-polA (**C**) in their respective

reaction buffer (**Table 2**). PCR program was (2 min at 94°C) × 1; (1 min at the indicated temperature, 1 min at 58°C, 2 min at 72°C) × 30; (5 min at 72°C) × 1. Molecular weight markers (M) are SmartLadder SF from Eurogentec.

by Pab-polB (**Figure 4B**). Pab-polD exhibited a lower extension rate than Taq-polA since a single specific product at the expected size appeared at 360 and 300 s, respectively for each enzyme. According to these results, the rate of extension by Pab-polD was estimated at 0.33 kb/min, resembling that of Taq-polA (0.39 kb/min), but dissimilar to that of Pab-polB (0.48 kb/min) (**Table 2**).

PCR AMPLIFICATION OF DNA FRAGMENTS OF VARIOUS LENGTHS

To determine the ability of Pab-polD to amplify various sized target sequences (ranging 0.5 to 10 kb) from genomic DNA, six specific primers have been designed (**Table 1**). The extension time assigned to the amplification of each DNA fragment during cycling is in agreement with the extension rate of Pab-polD described above. In the range of 0.5–1.1 kb, Pab-polD and Taq-polA efficiently and specifically amplified DNA fragments (**Figures 5A,C**). Amplification of DNA molecules ranging from 1.95 to 2.95 kb was severely reduced for Pab-polD and to lower extent for Taq-polA (~8% activity for the 2.95 kb target compared with the 0.5 kb target for both DNA polymerases), with bands at these sizes being faint and slightly detectable in ethidium bromide stained agarose gels after 30 PCR cycles. Amplification of target sequences ranging from 4.15 to 10 kb by Pab-polD and Taq-polA failed in the conditions tested (**Figures 5A,C**), with Pab-polD producing weak and unspecific bands. With the exception of the 4.15 and 10 kb DNA targets, Pab-polD amplified single products with the expected size (**Figure 5B**). Except for the 10 kb DNA target, a single product was amplified with Pab-polB. However, the yield of specific PCR products is decreased for the 2.95 kb DNA target. The above data clearly show that Pab-polD is suitable for the specific amplification

of DNA molecules in the range of 0.5–2.95 kb, while showing reduced yields above 1.1 kb. Although not shown, replacement of Pab-polD optimal buffer conditions by Taq-polA or Pab-polB reaction buffer did not improve the yield of amplified 1.95–2.95 kb DNA targets, nor enhanced the synthesis of longer PCR products.

PCR AMPLIFICATION IN THE PRESENCE OF PRIMER MISMATCHES

Complete 3'-terminal primer annealing to its complementary target sequence is a very important factor for the success and stringency of PCR (Petruska et al., 1988; Ishii and Fukui, 2001; Sipos et al., 2007). To evaluate the impact of primer mismatches on PCR efficiency and specificity by Pab-polD, forward primer sets containing up to four mismatches at the 3'-end have been designed for full amplification of the 0.5 kb DNA target from genomic DNA (**Table 1**). As shown in **Figure 6A**, specific amplification of the 0.5 kb DNA target could be achieved in the presence of either one or two 3'-end terminal mismatches, with lower PCR efficiency observed with two mismatches (~5% remaining activity). The presence of three or four mismatches had a detrimental effect on the extension efficiency by Pab-polD. Taq-polA DNA polymerase generated specific PCR products with only one mismatched primer termini and longer mismatches prevented successful PCR amplification (**Figure 6C**). Although strand extension, and hence PCR amplification efficiency, were influenced by multiple mismatches at the 3' end of the primer, the detection of specific PCR products was never compromised for Pab-polB (**Figure 6B**). Here, the data pointed out that Pab-polD is a suitable enzyme for Taq-polA substitution when 3'-terminal mismatched primers are refractory to PCR amplification.

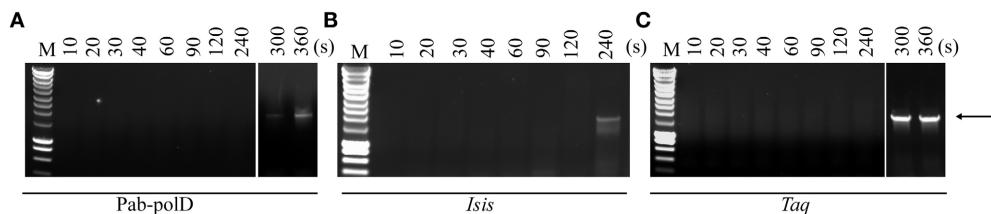


FIGURE 4 | Rate of DNA extension. PCR amplification of the 1.95 kb target (**Table 1**) was carried out with 0.1 U of Pab-polD (**A**), 1 U of Pab-polB (**B**), and 1 U of Taq-polA (**C**) in their respective reaction buffer (**Table 2**). PCR program was (2 min at 94°C) × 1; (1 min at 94°C, 1 min at 58°C, varying times in seconds as indicated at 72°C) × 30. Molecular weight markers (M) are SmartLadder LF from Eurogentec. The arrow indicates the specific 1.95 kb band.

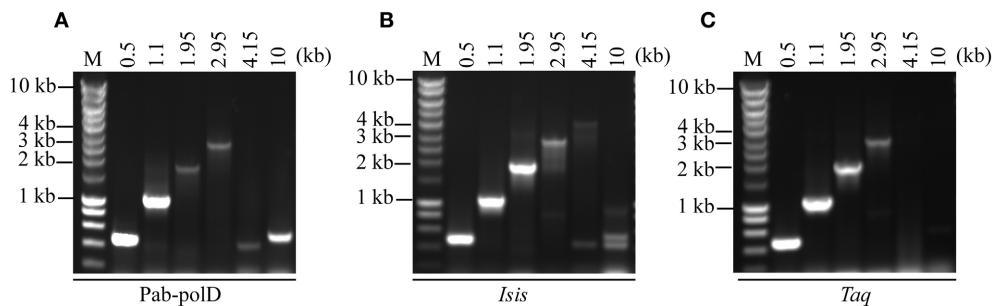


FIGURE 5 | PCR amplification of DNA fragments of various lengths. PCR reactions were carried out using 100 ng of genomic DNA with 0.1 U of Pab-polD (**A**), 1 U of Pab-polB (**B**), and 1 U of Taq-polA (**C**) in their respective reaction buffer (**Table 2**). Primer sets were chosen to amplify 0.5, 1.1, 1.95, 2.95, 4.15, and 10 kb (**Table 1**). PCR programs were (2 min at 94°C) × 1; (1 min at 94°C/1 min at 58°C/2, 4, 6, 8, 11, and 16 min with respect to the target length at 72°C) × 30. Molecular weight markers (M) are SmartLadder LF from Eurogentec.

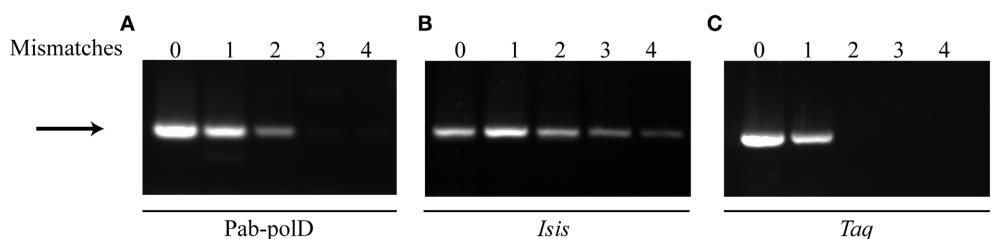


FIGURE 6 | PCR amplification in the presence of primer mismatches. PCR amplification of the 0.5 kb target (**Table 1**) was carried out with 0.1 U of Pab-polD (**A**), 1 U of Pab-polB (**B**), and 1 U of Taq-polA (**C**) in their respective reaction buffer (**Table 2**). Primer sets were chosen to

introduce 0, 1, 2, 3, and 4 mismatches at the 3'-termini of the forward primer (**Table 1**). PCR program was (2 min at 94°C) × 1; (1 min at 94°C, 1 min at 58°C, 2 min at 72°C) × 30; (5 min at 72°C) × 1. The arrow indicates the specific 0.5 kb band.

PCR AMPLIFICATION IN THE PRESENCE OF KNOWN PCR INHIBITORS

To investigate the impact of known PCR inhibitors (Al-Soud and Radstrom, 1998, 2001; Schrader et al., 2012), the 0.5 kb target was amplified from genomic DNA in reactions containing various levels of PCR-inhibiting compounds. The results are presented in **Table 3**. It was found that the variation of the ionic strength affected the PCR performances of the three different DNA polymerases. Indeed, Pab-polD was more resistant to NaCl ions than were Pab-polB and Taq-polA. In the conditions tested, a specific PCR product was detectable at the permissive concentration of 50 mM with Pab-polD, while absent with Pab-polB and Taq-polA (Supplementary Figure 1). When SDS (Sodium Dodecyl Sulfate), the anionic reagent well-known for its protein-denaturing effects, was employed at a concentration of 0.02%, PCR amplification was

successful with Pab-polD and Pab-polB (**Table 3**); however, the yield of PCR products was dramatically impaired (Supplementary Figure 1). In comparison, Taq-polA was weakly active at a SDS concentration of 0.01%. CaCl₂, a potent PCR inhibitor (Bickley et al., 1996; Al-Soud and Radstrom, 1998) found for instance in milk, cheese, or bones, was found to impede the amplification of the specific 0.5 kb target when incubated at a concentration of 2 mM for Pab-polD and Taq-polA (**Table 3**). Family B DNA polymerase was highly sensitive to CaCl₂ as observed the severe reduction in PCR activity at a final concentration of 1 mM (Supplementary Figure 1).

It has been reported that heme compounds, for instance hemoglobin and hematin, yielded interferences in PCR amplifications (Akane et al., 1994). For these reasons, we investigated

Pab-polD PCR performances in the presence of these two blood-substances. Interestingly, hemoglobin never altered PCR specificity and efficiency by Pab-PoD and Pab-polB even at a final concentration of 13 mg/ml (Supplementary Figure 1). However, PCR performances by Taq-polA were entirely compromised at the minimal concentrations of 1.75–3.5 mg/ml. Hematin decreased the yield of PCR products by Pab-polD and Taq-polA when present at a final concentration of 12.5 μM, while Pab-polB still retained significant activity in the presence of >25 μM of hematin.

The blood anticoagulant substance heparin described to inhibit PCR (Yokota et al., 1999) has been investigated. In the PCR reactions, the addition of heparin suppressed the formation of a specific PCR product in a dose-dependent fashion (from 0.006 to 0.2 U/μl) for all three PCR enzymes. Moreover, EDTA (EthyleneDiamineTetraAcetic acid), used also as a common anti-coagulant to treat blood samples, and included in several elution buffers of nucleic acid purification, has been described to interfere with PCR (Yokota et al., 1999). In this study, EDTA had an inhibitory effect at a concentration greater than 0.5 mM for all three PCR enzymes. This result is likely indicative of the chelation of Mg²⁺ ions present in reaction buffers of each DNA polymerase, therefore compromising DNA amplification.

Since urea has been identified as the main component of urine that inhibits PCR (Khan et al., 1991), we challenged Pab-polD, Pab-polB, and Taq-polA in its presence. As shown in **Table 3**, Pab-polB was the most resistant DNA polymerase, specifically amplifying the 0.5 kb target even at the highest concentration of 100 mM. Pab-polD and Taq-polA showed inhibitory concentrations >25 and 50 mM, respectively. Humic acid, representative of environmental samples (e.g., soil, water, and dead matter), has been recognized as an efficient PCR inhibitor even at low concentrations (Tsai and Olson, 1992; Ijzerman et al., 1997). For this reason, all three enzymes have been submitted to PCR amplification with increasing amounts of humic acid (15–250 ng/μl). No PCR product was visible with Pab-polD, even at the lowest

concentration. PCR amplification with Taq-polA was positive up to a humic acid concentration of 15 ng/μl. Pab-polB was the most resistant since it could PCR amplify the 0.5 kb target at 62.5 ng/μl of humic acid (**Table 3**).

Phenol, ethanol, and isopropanol are common organic substances used in the procedure for genomic extraction (Charbonnier et al., 1995) and relevant of food and environmental samples (Wilson, 1997). Interestingly, PCR performances of family B DNA polymerase were never affected by the varying concentrations of the three organic compounds (Supplementary Figure 1). In contrast, Taq-polA and Pab-polD exhibited similar inhibitory concentrations of isopropanol and ethanol (>2 and >4%, respectively). Moreover, Pab-polD was more tolerant to PCR inhibition by phenol than Taq-polA (**Table 3**). Overall, this comparative study clearly revealed that Pab-polB is the most tolerant enzyme to PCR inhibitors. Pab-polD seems to have a higher resistance to particular PCR inhibitors than Taq-polA (e.g., NaCl, SDS, Phenol, and hemoglobin), although sharing similar sensitivity (EDTA, ethanol, isopropanol, calcium, hematin, heparin). Finally, Taq-polA exhibited superior resistance to humic acid and urea than Pab-polD.

DISCUSSION

The family D DNA polymerase from *P. abyssi* has been applied to PCR on genomic DNA and submitted to varying chemical parameters in order to evaluate its performance. For this purpose, a buffer has been optimized using the Pab-polB reaction buffer as a starting point due to this family B DNA polymerase originated from the same archaeon, *P. abyssi*. Here, we show for the first time that a family D DNA polymerase is functional in PCR amplification of a 0.5 kb DNA target. Under the conditions listed in **Table 1**, the three enzymes were analyzed to compare thermal resistance. Pab-polD was identified as more resistant than Taq-polA, yet not as robust as Pab-polB to the increased thermal denaturation during cycling (**Table 2**). These data obtained for the commercial enzymes being comparable with that already published (Gueguen et al., 2001; Dietrich et al., 2002) (Supplementary Figure 2) clearly suggest that Pab-polD is a thermostable enzyme. These results are interesting since they indicate that denaturation temperature during cycling can be increased when required during PCR. This is particularly useful when genomic DNA contains secondary structures or high GC-rich regions.

As expected, for all three enzymes PCR efficiency was variable in respect to the concentration of the 1.7 million base-pair genomic DNA, with 0.5 ng being the permissive amount for Pab-polB, while Pab-polD and Taq-polA were more sensitive to template dilution. PCR analysis of trace amounts of DNA has become an important concern in forensic investigations (Van Oorschot et al., 2010). While some laboratories set up 0.2 ng as a threshold limit for reliability of the investigations (Budowle et al., 2009), others continue to revise this limit (Kaminiwa et al., 2013). In our conditions, Pab-polD did not behave as an effective tool for the amplification of limited amounts of genomic DNA. Therefore, increasing the number of cycles or varying some compounds within the reaction buffer can be a useful alternative to overcome the limits (Van Oorschot et al., 2010).

Table 3 | Potential inhibitory effects of organic and inorganic substances on PCR.

	Inhibitory concentration of compounds		
	Pab-polD	Pab-polB	Taq-polA
NaCl (mM)	>50	>25	>25
SDS % (w/v)	>0.02	>0.02	>0.01
CaCl ₂ (mM)	>2	>1	>2
Hemoglobin (mg/ml)	>13	>13	>3.5
Hematin (μM)	>12.5	>25	>12.5
Heparin (U/ml)	–	–	–
EDTA (mM)	>0.5	>0.5	>0.5
Urea (mM)	>25	>100	>50
Humic acid (ng/μl)	–	>62	>15
Phenol % (v/v)	>0.6	>0.8	>0.4
Ethanol % (v/v)	>4	>5	>4
Isopropanol % (v/v)	>2	>8	>2

In this study, the highest amount of genomic DNA has been applied to all PCR experiments and in these conditions the 0.5 kb target was significantly amplified by Pab-polD, Pab-polB, and Taq-polA. However, upon increasing the length of the target, the differences in PCR performance became more obvious (Pab-polB > Taq-polA > Pab-polD). A maximum of length of 2.95 kb was produced by Taq-polA and Pab-polD. Although barely detectable, Pab-polB could amplify the 4.15 kb target. The difficulty of Pab-polD to PCR amplify long DNA fragments was not due to a high GC content of the DNA regions since all exhibited a value below 48% (**Table 1**). Pab-polD is known to be endowed with lower processivity than Pab-polB, requiring the PCNA (Proliferating Cell Nuclear Antigen) clamp for robust DNA synthesis (Henneke et al., 2005). Thus, further optimization of PCR amplification of large DNA fragments is certainly possible, for instance, by altering the reaction buffer components, adding PCNA or mixing the two *Pab* PCR enzymes.

Full annealing between primer and template sequences is generally considered crucial for the specific amplification of a nucleic acid sequence (Ghadessy et al., 2004). PCR-based amplification of specific sequences is essential in detecting single nucleotide polymorphisms (SNPs), in identifying microbial-archaeal populations and in diagnostics (Sipos et al., 2007; Liu et al., 2012). In these approaches, “universal” primer sets are used with the possibility to induce the formation of mismatched base pairs at template-primer 3'-termini. As a result, PCR amplification is reduced or fully inhibited (Huang et al., 1992), depending on the length of base mispairs. In our study, Pab-polD was challenged in PCR with mismatched base pairs (1, 2, 3, and 4 base mispairs) at template-primer 3'-termini. Pab-polD could amplify the 0.5 kb DNA target despite the presence of 2 mismatches but with reduced efficiency. Taq-polA retains activity in the presence of 1 mismatch at the 3'-terminus as already published (Huang et al., 1992) but was inhibited by two mismatches. Pab-polB was functional even in the presence of 4 mismatches. The *Pab*PCR enzymes show a higher tolerance to the presence of mismatches which must be attributed to their associated 3'-5' exonuclease function as already compared (Gueguen et al., 2001). Up to four and two mismatches can be accommodated into the exonuclease active site of Pab-polB and Pab-polD, respectively, which subsequently induce the degradation of the 3'-termini. Families B and D from *Pfu* are also known to efficiently process 3'-end termini of primers (Richardson et al., 2013a,b).

Time-dependent PCR extension has been carried out with Pab-polD and compared to Pab-polB and Taq-polA. A PCR product of constant length (1.95 kb) was amplified by increasing the extension time. Under these conditions, Pab-polD was the slowest enzyme able to generate the 1.95 kb target in 6 min whereas 5 and 4 min were required respectively for Taq-polA and Pab-polB. The calculated extension rate of Pab-polD was 0.33 kb/min. This is almost comparable with the value of 0.39 kb/min for Taq-polA and slightly lower to that of 0.48 kb/min for Pab-polB. The values confirmed that Pab-polB, like other family B DNA polymerases, e.g., *Pfu* and *Tfu*, are particularly slow enzymes (Perler et al., 1996; Terpe, 2013). Although the extension rate has been determined by conventional end-point PCR which is not the method of choice compared to real-time quantitative PCR (Arezi

et al., 2003), Pab-polD also shows a reduced elongation rate in PCR. In Pab-polB, and potentially Pab-polD, this property could be explained by the slow kinetic partitioning of the primer in the exonuclease and polymerase active sites allowing the DNA polymerase to proofread the nucleotide incorporation events, and when required to remove the misincorporated base (Gouge et al., 2012). On the other hand, the presence of secondary structures could also impede the efficiency of DNA synthesis by the *Pyrococcus* enzymes (Henneke, 2012).

The negative effect of inorganic and organic substances on PCR efficiency and specificity by Pab-polD along with Pab-polB and Taq-polA has been investigated. The DNA polymerase the most resistant to ions NaCl and CaCl₂ was Pab-polD. This higher resistance to elevated NaCl concentrations is similar to that found for some bacterial and archaeal PCR enzymes investigated previously (Al-Soud and Radstrom, 1998). The highest tolerance to calcium ions compared to other thermostable DNA polymerases (Al-Soud and Radstrom, 1998) places Pab-polD as a suitable enzyme in the amplification of food (e.g., milk and cheese) and human samples (e.g., teeth and bones) for instance. Introduced during the procedure of genomic extraction or naturally present in food and in environmental samples (Charbonnier et al., 1995; Wilson, 1997), phenolic compounds (Ethanol, phenol, and isopropanol) reduced the PCR performance of Pab-polD. These negative effects are commonly observed with most PCR enzymes (Rossen et al., 1992), except for Pab-polB (shown in this study) and *Tth* (*Thermus thermophilus*) (Katcher and Schwartz, 1994). Compounds, such as the SDS anionic detergent and EDTA, known to have direct and indirect negative effects on proteins, respectively, did not dramatically impact on PCR performances by Pab-polD compared to Pab-polB. Surprisingly, the permissive concentration of the two compounds for Taq-polA were slightly different to those found in another study (Yang et al., 2007), indicating that the source of the enzyme, the DNA target to be amplified and the reaction conditions are important parameters impacting on the issue of the investigation. The inhibitory activity of urea in PCR was observed with Pab-polD at a lower concentration threshold compared to Taq-polA or Pab-polB, and the values obtained with Taq-polA confirmed those previously published (Khan et al., 1991). Organic substances like heparin or humic acid were completely inhibitory to PCR reactions by Pab-polD. The strong inhibitory effect of heparin on all three PCR enzymes has been observed with other thermostable DNA polymerases (Yokota et al., 1999). This is not so surprising since heparin is commonly used as a trapping agent of DNA polymerases in both polymerase assays and chromatography. The effects of two heme blood substances, hemoglobin and hematin, did not impact similarly on PCR performances by Pab-polD. While completely resistant to hemoglobin, Pab-polD was sensitive to hematin but to the same level as Taq-polA. Generally, inhibitory effects by heme compounds appear as a drawback in PCR with thresholds dependent on the PCR enzymes used (Akane et al., 1994; Al-Soud and Radstrom, 2001). In this study, Taq-polA was the most sensitive to hemoglobin.

In conclusion, our results demonstrated for the first time that an archaeal family D DNA polymerase is functional in PCR. PCR performances (rate of DNA synthesis, maximal length of

amplification, minimal input genomic DNA, resistance to thermal denaturation during cycling, and PCR amplifications with 3'-end mismatched primers) of Pab-polD appears more comparable to Taq-polA than Pab-polB, but with some valuable properties (e.g., high resistance to thermal denaturation during cycling, amplification with primers containing up to 2 mismatches). In addition, due to its superior resistance to inhibitors than Taq-polA (e.g., calcium ions, sodium chloride, hemoglobin, SDS), Pab-polD could replace the enzyme in some applications. Additional investigations (e.g., PCR fidelity, ability to specifically amplify high GC rich content and degraded genomic DNA) are now required to consider Pab-polD as a suitable PCR enzyme that could overcome the handicap encounter by conventional enzymes that are marketed for PCR.

ACKNOWLEDGMENTS

The work was financially supported by the French National Research Agency (ANR-10-JCJC-1501-01 to Ghislaine Henneke) and the General Council of Finistère (CG29). I would like to thank Sébastien Le Laz for help with PCR optimization.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00195/abstract>

REFERENCES

- Akane, A., Matsubara, K., Nakamura, H., Takahashi, S., and Kimura, K. (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J. Forensic Sci.* 39, 362–372.
- Al-Soud, W. A., and Radstrom, P. (1998). Capacity of nine thermostable DNA polymerases To mediate DNA amplification in the presence of PCR-inhibiting samples. *Appl. Environ. Microbiol.* 64, 3748–3753.
- Al-Soud, W. A., and Radstrom, P. (2001). Purification and characterization of PCR-inhibitory components in blood cells. *J. Clin. Microbiol.* 39, 485–493. doi: 10.1128/JCM.39.2.485-493.2001
- Arezi, B., Xing, W., Sorge, J. A., and Hogrefe, H. H. (2003). Amplification efficiency of thermostable DNA polymerases. *Anal. Biochem.* 321, 226–235. doi: 10.1016/S0003-2697(03)00465-2
- Bickley, J., Short, J. K., McDowell, D. G., and Parkes, H. C. (1996). Polymerase chain reaction (PCR) detection of *Listeria monocytogenes* in diluted milk and reversal of PCR inhibition caused by calcium ions. *Lett. Appl. Microbiol.* 22, 153–158. doi: 10.1111/j.1472-765X.1996.tb01131
- Biles, B. D., and Connolly, B. A. (2004). Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res.* 32, e176. doi: 10.1093/nar/gnh174
- Braithwaite, D. K., and Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21, 787–802. doi: 10.1093/nar/21.4.787
- Budowle, B., Eisenberg, A. J., and Van Daal, A. (2009). Validity of low copy number typing and applications to forensic science. *Croat. Med. J.* 50, 207–217. doi: 10.3325/cmj.2009.50.20.207
- Cambon-Bonavita, M. A., Schmitt, P., Zieger, M., Flaman, J. M., Lesongeur, F., Ragueneau, G., et al. (2000). Cloning, expression, and characterization of DNA polymerase I from the hyperthermophilic archaea *Thermococcus fumicola*. *Extremophiles* 4, 215–225. doi: 10.1007/PL00010714
- Cann, I. K., and Ishino, Y. (1999). Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* 152, 1249–1267.
- Cann, I. K., Komori, K., Toh, H., Kanai, S., and Ishino, Y. (1998). A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14250–14255. doi: 10.1073/pnas.95.24.14250
- Castrec, B., Laurent, S., Henneke, G., Flament, D., and Raffin, J. P. (2010). The glycine-rich motif of *Pyrococcus abyssi* DNA polymerase D is critical for protein stability. *J. Mol. Biol.* 396, 840–848. doi: 10.1016/j.jmb.2010.01.006
- Castrec, B., Rouillon, C., Henneke, G., Flament, D., Querellou, J., and Raffin, J. P. (2009). Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. *J. Mol. Biol.* 394, 209–218. doi: 10.1016/j.jmb.2009.09.044
- Charbonnier, F., Forterre, P., Erauso, G., and Prieur, D. (1995). “Purification of plasmids from thermophilic and hyperthermophilic archaea,” in *Archaea: A Laboratory Manual*, eds A. R. Place and F. T. Robb (New York, NY: Cold Spring Harbor Laboratory Press), 87–90.
- Chien, A., Edgar, D. B., and Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* 127, 1550–1557.
- Cline, J., Braman, J. C., and Hogrefe, H. H. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* 24, 3546–3551. doi: 10.1093/nar/24.18.3546
- Cubanova, L., Richardson, T., Burkhardt, B. W., Kelman, Z., Connolly, B. A., Reeve, J. N., et al. (2013). Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *J. Bacteriol.* 195, 2322–2328. doi: 10.1128/JB.02037-12
- Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* 3, 461–467. doi: 10.1093/protein/3.6.461
- Dietrich, J., Schmitt, P., Zieger, M., Preve, B., Rolland, J. L., Chaabibi, H., et al. (2002). PCR performance of the highly thermostable proof-reading B-type DNA polymerase from *Pyrococcus abyssi*. *FEMS Microbiol. Lett.* 217, 89–94. doi: 10.1016/S0378-1097(02)01037-6
- Eckert, K. A., and Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl.* 1, 17–24.
- Fogg, M. J., Pearl, L. H., and Connolly, B. A. (2002). Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* 9, 922–927. doi: 10.1038/nsb867
- Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., et al. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* 22, 755–759. doi: 10.1038/nbt974
- Gouge, J., Ralec, C., Henneke, G., and Delarue, M. (2012). Molecular recognition of canonical and deaminated bases by *P. abyssi* family B DNA polymerase. *J. Mol. Biol.* 423, 315–336. doi: 10.1016/j.jmb.2012.07.025
- Gueguen, Y., Rolland, J. L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., et al. (2001). Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur. J. Biochem.* 268, 5961–5969. doi: 10.1046/j.0014-2956.2001.02550.x
- Henneke, G. (2012). *In vitro* reconstitution of RNA primer removal in Archaea reveals the existence of two pathways. *Biochem. J.* 447, 271–280. doi: 10.1042/BJ20120959
- Henneke, G., Flament, D., Hubscher, U., Querellou, J., and Raffin, J. P. (2005). The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J. Mol. Biol.* 350, 53–64. doi: 10.1016/j.jmb.2005.04.042
- Hogrefe, H. H., Cline, J., Lovejoy, A. E., and Nielson, K. B. (2001). DNA polymerases from hyperthermophiles. *Methods Enzymol.* 334, 91–116. doi: 10.1016/S0076-6879(01)34461-0
- Huang, M. M., Arnheim, N., and Goodman, M. F. (1992). Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* 20, 4567–4573. doi: 10.1093/nar/20.17.4567
- Hubscher, U., Spadari, S., Villani, G., and Maga, G. (2010). *DNA Polymerases: Discovery, Characterization and Functions in Cellular DNA Transactions*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Ijzerman, M. M., Dahling, D. R., and Fout, G. S. (1997). A method to remove environmental inhibitors prior to the detection of waterborne enteric viruses by reverse transcription-polymerase chain reaction. *J. Virol. Methods* 63, 145–153. doi: 10.1016/S0166-0934(96)02123-4
- Ishii, K., and Fukui, M. (2001). Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl. Environ. Microbiol.* 67, 3753–3755. doi: 10.1128/AEM.67.8.3753-3755.2001
- Ishino, Y., Komori, K., Cann, I. K., and Koga, Y. (1998). A novel DNA polymerase family found in Archaea. *J. Bacteriol.* 180, 2232–2236.

- Joyce, C. M., and Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63, 777–822. doi: 10.1146/annurev.bi.63.070194.004021
- Kaminiwa, J., Honda, K., Sugano, Y., Yano, S., Nishi, T., and Sekine, Y. (2013). Vanadium accelerates polymerase chain reaction and expands the applicability of forensic DNA testing. *J. Forensic Leg. Med.* 20, 326–333. doi: 10.1016/j.jflm.2012.09.006.
- Katcher, H. L., and Schwartz, I. (1994). A distinctive property of Tth DNA polymerase: enzymatic amplification in the presence of phenol. *Biotechniques* 16, 84–92.
- Khan, G., Kangro, H. O., Coates, P. J., and Heath, R. B. (1991). Inhibitory effects of urine on the polymerase chain reaction for cytomegalovirus DNA. *J. Clin. Pathol.* 44, 360–365. doi: 10.1136/jcp.44.5.360
- Kong, H., Kucera, R. B., and Jack, W. E. (1993). Characterization of a DNA polymerase from the hyperthermophilic archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J. Biol. Chem.* 268, 1965–1975.
- Kornberg, A., and Baker, T. A. (1992). *DNA Replication*. New York, NY: W. H. Freeman and Company.
- Langhorst, B. W., Jack, W. E., Reha-Krantz, L., and Nichols, N. M. (2012). Polbase: a repository of biochemical, genetic and structural information about DNA polymerases. *Nucleic Acids Res.* 40, D381–387. doi: 10.1093/nar/gkr847
- Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., et al. (1993). High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl.* 2, 275–287.
- Le, Y., Chen, H., Zagursky, R., Wu, J. H., and Shao, W. (2013). Thermostable DNA ligase-mediated PCR production of circular plasmid (PPCP) and its application in directed evolution via *in situ* error-prone PCR. *DNA Res.* 20, 375–382. doi: 10.1093/dnares/dst016
- Lipps, G., Rothe, S., Hart, C., and Krauss, G. (2003). A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. *EMBO J.* 22, 2516–2525. doi: 10.1093/emboj/cdg246
- Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., et al. (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* 8, 34. doi: 10.1186/1746-4811-8-34
- Macneill, S. A., Baldacci, G., Burgers, P. M., and Hubscher, U. (2001). A unified nomenclature for the subunits of eukaryotic DNA polymerase delta. *Trends Biochem. Sci.* 26, 16–17. doi: 10.1016/S0968-0004(00)01709-6
- Marsic, D., Flaman, J. M., and Ng, J. D. (2008). New DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus thioreducens*. *Extremophiles* 12, 775–788. doi: 10.1007/s00792-008-0181-7
- Matsui, I., Urushibata, Y., Shen, Y., Matsui, E., and Yokoyama, H. (2011). Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from *Pyrococcus horikoshii*. *FEBS Lett.* 585, 452–458. doi: 10.1016/j.febslet.2010.12.040
- McCullum, E. O., Williams, B. A., Zhang, J., and Chaput, J. C. (2010). Random mutagenesis by error-prone PCR. *Methods Mol. Biol.* 634, 103–109. doi: 10.1007/978-1-60761-652-8_7
- McDonald, J. P., Hall, A., Gasparutto, D., Cadet, J., Ballantyne, J., and Woodgate, R. (2006). Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res.* 34, 1102–1111. doi: 10.1093/nar/gkj512
- Moussard, H., Henneke, G., Moreira, D., Jouffe, V., Lopez-Garcia, P., and Jeanthon, C. (2006). Thermophilic lifestyle for an uncultured archaeon from hydrothermal vents: evidence from environmental genomics. *Appl. Environ. Microbiol.* 72, 2268–2271. doi: 10.1128/AEM.72.3.2268-2271.2006
- Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., et al. (2001). The Y-family of DNA polymerases. *Mol. Cell* 8, 7–8. doi: 10.1016/S1097-2765(01)00278-7
- Palud, A., Villani, G., L'Haridon, S., Querellou, J., Raffin, J. P., and Henneke, G. (2008). Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance? *Mol. Microbiol.* 70, 746–761. doi: 10.1111/j.1365-2958.2008.06446.x
- Pavlov, A. R., Pavlova, N. V., Kozyavkin, S. A., and Slesarev, A. I. (2004). Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol.* 22, 253–260. doi: 10.1016/j.tibtech.2004.02.011
- Perler, F. B., Kumar, S., and Kong, H. (1996). Thermostable DNA polymerases. *Adv. Protein Chem.* 48, 377–435. doi: 10.1016/S0065-3233(08)60367-8
- Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C., and Tinoco, I. Jr. (1988). Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc. Natl. Acad. Sci. U.S.A.* 85, 6252–6256. doi: 10.1073/pnas.85.17.6252
- Richardson, T. T., Gilroy, L., Ishino, Y., Connolly, B. A., and Henneke, G. (2013a). Novel inhibition of archaeal family-D DNA polymerase by uracil. *Nucleic Acids Res.* 41, 4207–4218. doi: 10.1093/nar/gkt083
- Richardson, T. T., Wu, X., Keith, B. J., Heslop, P., Jones, A. C., and Connolly, B. A. (2013b). Unwinding of primer-templates by archaeal family-B DNA polymerases in response to template-strand uracil. *Nucleic Acids Res.* 41, 2466–2478. doi: 10.1093/nar/gks1364
- Rossen, L., Norskov, P., Holmstrom, K., and Rasmussen, O. F. (1992). Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. *Int. J. Food Microbiol.* 17, 37–45. doi: 10.1016/0168-1605(92)90017-W
- Rothwell, P. J., and Waksman, G. (2005). Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* 71, 401–440. doi: 10.1016/S0065-3233(04)71011-6
- Rouillon, C., Henneke, G., Flament, D., Querellou, J., and Raffin, J. P. (2007). DNA polymerase switching on homotrimeric PCNA at the replication fork of the euryarchaea *Pyrococcus abyssi*. *J. Mol. Biol.* 369, 343–355. doi: 10.1016/j.jmb.2007.03.054
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354. doi: 10.1126/science.299980
- Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors – occurrence, properties and removal. *J. Appl. Microbiol.* 113, 1014–1026. doi: 10.1111/j.1365-2672.2012.05384.x
- Sharkey, D. J., Scalise, E. R., Christy, K. G. Jr., Atwood, S. M., and Daiss, J. L. (1994). Antibodies as thermolabile switches: high temperature triggering for the polymerase chain reaction. *Biotechnology (N.Y.)* 12, 506–509. doi: 10.1038/nbt0594-506
- Shen, Y., Musti, K., Hiramoto, M., Kikuchi, H., Kawarabayashi, Y., and Matsui, I. (2001). Invariant Asp-1122 and Asp-1124 are essential residues for polymerization catalysis of family D DNA polymerase from *Pyrococcus horikoshii*. *J. Biol. Chem.* 276, 27376–27383. doi: 10.1074/jbc.M011762200
- Sipos, R., Szekely, A. J., Palatinszky, M., Revesz, S., Mariáligeti, K., and Niklausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol. Ecol.* 60, 341–350. doi: 10.1111/j.1574-6941.2007.00283.x
- Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., et al. (1997). Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* 63, 4504–4510.
- Terpe, K. (2013). Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 97, 10243–10254. doi: 10.1007/s00253-013-5290-2
- Tsai, Y. L., and Olson, B. H. (1992). Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl. Environ. Microbiol.* 58, 2292–2295.
- Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, Y. (1997). A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes Cells* 2, 499–512. doi: 10.1046/j.1365-2448.1997.1380336.x
- Van Oorschot, R. A., Ballantyne, K. N., and Mitchell, R. J. (2010). Forensic trace DNA: a review. *Investig. Genet.* 1, 14. doi: 10.1186/2041-2223-1-14
- Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* 32, 1197–1207. doi: 10.1093/nar/gkh271
- Wilson, I. G. (1997). Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* 63, 3741–3751.

- Wynne, S. A., Pinheiro, V. B., Holliger, P., and Leslie, A. G. (2013). Structures of an apo and a binary complex of an evolved archaeal B family DNA polymerase capable of synthesising highly cy-dye labelled DNA. *PLoS ONE* 8:e70892. doi: 10.1371/journal.pone.0070892
- Yamasaki, K., Urushibata, Y., Yamasaki, T., Arisaka, F., and Matsui, I. (2010). Solution structure of the N-terminal domain of the archaeal D-family DNA polymerase small subunit reveals evolutionary relationship to eukaryotic B-family polymerases. *FEBS Lett.* 584, 3370–3375. doi: 10.1016/j.febslet.2010.06.026
- Yang, Y. G., Kim, J. Y., Song, Y. H., and Kim, D. S. (2007). A novel buffer system, AnyDirect, can improve polymerase chain reaction from whole blood without DNA isolation. *Clin. Chim. Acta* 380, 112–117. doi: 10.1016/j.cca.2007.01.019
- Yokota, M., Tatsumi, N., Nathalang, O., Yamada, T., and Tsuda, I. (1999). Effects of heparin on polymerase chain reaction for blood white cells. *J. Clin. Lab. Anal.* 13, 133–140. doi: 10.1002/(SICI)1098-2825(1999)13:3<133::AID-JCLA8>3.0.CO;2-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 March 2014; accepted: 12 April 2014; published online: 07 May 2014.

Citation: Killelea T, Ralec C, Bosse A and Henneke G (2014) PCR performance of a thermostable heterodimeric archaeal DNA polymerase. *Front. Microbiol.* 5:195. doi: 10.3389/fmicb.2014.00195

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Killelea, Ralec, Bosse and Henneke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Compartmentalized self-replication under fast PCR cycling conditions yields *Taq* DNA polymerase mutants with increased DNA-binding affinity and blood resistance

Bahram Arezi*, Nancy McKinney, Connie Hansen, Michelle Cayouette, Jeffrey Fox, Keith Chen, Jennifer Lapira, Sarah Hamilton and Holly Hogrefe

Agilent Technologies, La Jolla, CA, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Phil Holliger, MRC Laboratory of Molecular Biology, UK
Jennifer Lee Ong, New England Biolabs, USA

***Correspondence:**

Bahram Arezi, Agilent Technologies, 11011 North Torrey Pines Rd., La Jolla, CA 92037, USA
e-mail: bahram.arezi@agilent.com

Faster-cycling PCR formulations, protocols, and instruments have been developed to address the need for increased throughput and shorter turn-around times for PCR-based assays. Although run times can be cut by up to 50%, shorter cycle times have been correlated with lower detection sensitivity and increased variability. To address these concerns, we applied Compartmentalized Self Replication (CSR) to evolve faster-cycling mutants of *Taq* DNA polymerase. After five rounds of selection using progressively shorter PCR extension times, individual mutations identified in the fastest-cycling clones were randomly combined using ligation-based multi-site mutagenesis. The best-performing combinatorial mutants exhibit 35- to 90-fold higher affinity (lower K_d) for primed template and a moderate (2-fold) increase in extension rate compared to wild-type *Taq*. Further characterization revealed that CSR-selected mutations provide increased resistance to inhibitors, and most notably, enable direct amplification from up to 65% whole blood. We discuss the contribution of individual mutations to fast-cycling and blood-resistant phenotypes.

Keywords: fast PCR, fast cycling, *Taq* mutants, blood resistant, inhibitor resistant

INTRODUCTION

Taq DNA polymerase is still considered the workhorse of PCR, providing great economy and reliability in routine amplification of genomic targets up to 2 kb. Moreover, PCR detection in real-time relies on *Taq*'s intrinsic 5'-flap endonuclease activity for TaqMan probe hydrolysis and lack of proofreading activity to minimize primer/probe degradation (Holland et al., 1991). With an extension rate of 60 nt/s, wild-type *Taq* produces high amplicon yields after 30–40 cycles using 1 min anneal-extension times. Run times are 1.5–2 h on conventional Peltier-based PCR instruments and approximately 1 h using advanced qPCR instrumentation with improved thermal ramp rates (2.2–3°C/s). Demand for higher throughput and shorter turn-around-time continues to fuel interest in developing faster PCR instrumentation, along with polymerases with improved kinetic properties. In the near future, microchip-based technologies are expected to provide drastically reduced run times (<3–5 min), limited only by the kinetics of nucleotide incorporation (Hashimoto et al., 2004).

Currently, shorter qPCR run times are achieved by reducing hold times for denaturation, annealing and extension steps, and/or by using a 2-step cycling regimen with combined annealing and extension steps (at 60°C). Although run times can be cut by up to 50%, shorter cycle times with *Taq* have been correlated with lower detection sensitivity and higher failure rates when applied across a range of primer-template combinations (Hilscher et al., 2005). Results can be further improved by increasing the amount of *Taq* or other reagents, or by reducing reaction

volume and using thin-walled PCR tubes to further improve heat transfer. Ultimately, however, PCR run times are limited by the kinetic properties of the PCR enzyme. In the most telling example, the processivity of a proofreading Family B DNA polymerase was directly correlated to PCR cycle times. When the processivity of *Pfu* DNA polymerase was increased by 9-fold by fusion to a small basic double-stranded DNA-binding protein (Sso7d), PCR annealing/extension times could be reduced from 2 min to 30 s for a 5 kb λ target (Wang et al., 2004). Unfortunately, this strategy could not be employed to accelerate TaqMan assays, as *Pfu* lacks 5' endonuclease activity and Sso7d fusions to full-length *Taq* are unstable (data not shown).

To accelerate qPCR run times, we applied the Compartmentalized Self Replication (CSR) technique (Ghadessy et al., 2001) to evolve faster-cycling *Taq* mutants. CSR employs emulsion PCR to trap individual *E. coli* cells harboring mutant polymerase genes in microscopic aqueous compartments along with nucleotides and *pol* gene-specific primers. When the emulsion is cycled under selective conditions, active mutant polymerases self-replicate and are enriched, while those with insufficient activity fail to replicate and are lost from the gene pool. CSR has been used to successfully evolve *Taq* mutants with increased thermostability or heparin resistance (Ghadessy et al., 2001), and chimeric polymerases that are broadly resistant to complex environmental inhibitors or can process non-canonical primer-template duplexes and bypass lesions found in ancient DNA such as abasic sites (d'Abbadie et al., 2007). In this study, we used CSR

to evolve *Taq* mutants that can self-replicate under progressively shorter extension times. As we will show, CSR selection netted *Taq* variants with a broad range of beneficial attributes, in addition to increased polymerization rate.

MATERIALS AND METHODS

All molecular biology reagents were from Agilent Technologies unless otherwise noted. Oligonucleotides were purchased from Integrated DNA Technologies. Radioactive nucleotides [$\gamma^{33}\text{P}$] ATP-3000 Ci/mmol-1 mCi (NEG302H001MC) and deoxythymidine-5'-triphosphate [Methyl- ^3H] tetrasodium salt-1 mCi (NET221A001MC) were purchased from Perkin Elmer.

RANDOM AND SITE-DIRECTED MUTAGENESIS

Taq mutants were generated by random mutagenesis of the *Thermus aquaticus pol I* gene using the GeneMorph II random mutagenesis kit and PCR primers that introduce *Xba*I and *Sall* sites and an N-terminal His₆ tag (F: GGC GGCTCTAGATAACGAGGGCAAAAAATGCA TCATCA TCACC ATCAC, R: GCGGTGCGGAGTCGACTTACTCCTTGGC GGAGAGGCCAGTC). PCRs also included 5% DMSO and increasing amounts of plasmid template (10 ng, 1 ng, 0.1 ng) to generate libraries with varying mutation rates of 4.7–6.2 per kb. After *Dpn*I treatment, PCR products were gel purified and digested with *Xba*I and *Sall* (NEB). Purified fragments were cloned into the pASK-IBA5C expression vector (IBA) and transformed into XL10-Gold Kan cells. Site-directed mutagenesis was performed using the QuikChange Lightning or QuikChange Lightning Multi Site-Directed Mutagenesis kit.

CSR SELECTION

Approximately 250,000 independent clones were scraped from plates, re-suspended, and stored as glycerol stocks. LB/CAM cultures (40 ml) were freshly inoculated, grown at 30°C to OD₆₀₀ of 0.6, and induced with 200 ng/ml anhydrotetracycline. After 3 h of growth, cells were harvested and washed in 1× *Taq* buffer. CSR was carried out essentially as described by Ghadessy et al. (2001). To select for faster-cycling mutants, extension times were successively reduced over five rounds of CSR from 2.5 min (round 1) to 15 s (round 5). PCR selection was performed on a Robocycler 96 using 2.5 min at 94°C followed by 30 cycles of 45 s at 94°C, 30 s at 60°C, and 0.25–2.5 min at 72°C.

PROTEIN EXPRESSION

Colonies expressing mutant *Taq* polymerases were randomly picked, replicated, and then grown overnight at 30°C in 96-deep well plates (VWR) containing 750 μl LB/CAM. Overnight cultures (30 μl) were inoculated into fresh media, induced with anhydrotetracycline at OD_{600nm} of 0.3–0.5, and grown overnight with shaking at 30°C. Cells were collected and used to prepare lysates for direct PCR screening or for affinity protein purification (see below). Cell pellets were re-suspended in 50 μl Tris pH 8 containing 4 mg/ml lysozyme, and incubated at 37°C for 10 min to disrupt cell walls and at 75°C for 15 min to inactivate *E. coli* protein. Lysates were clarified by centrifugation for 15–30 min at 4000 RPM through a 96-well filter plate (Millipore, Multiscreen HTS, HV).

PURIFICATION OF HIS-TAGGED *Taq* MUTANTS

Cell pellets were re-suspended in 90 μl of buffer prepared by adding one cComplete EDTA-free Protease Inhibitor tablet (Roche) to 50 ml of 50 mM Tris pH 8, 0.5 M NaCl, 5 mM imidazole. An aliquot (60 μl) of 1× FastBreak Cell Lysis Reagent (Promega) was added and the lysate was incubated at 37°C for 15 min and at 70°C for 15 min, before centrifugation through a Millipore 96-well filter plate. Clarified lysates were combined with 60 μl Ni-NTA agarose (Qiagen) and incubated with shaking at room temperature for 2 h. After collecting the agarose resins using a fresh filter plate, resins were washed two times with 200 μl wash buffer (50 mM Tris pH 8, 0.5 M NaCl, 20 mM imidazole) and eluted with 80 μl of 50 mM Tris pH 8, 0.5 M NaCl, 200 mM imidazole.

FAST PCR SCREENING

Taq mutants were screened by amplifying a 549 bp GAPDH target on the Mx3005 qPCR system using fast-cycling conditions consisting of 1 min at 95°C followed by 50 cycles of 2 s at 99°C, 7 s at 59°C. PCRs (25 μl) contained 1× *Taq* buffer (15 mM Tris pH 8.0, 50 mM KCl, 2.5 mM MgCl₂, 0.01% Tween-20), 0.8 mM dNTPs, GAPDH primers (5'-ATCTTGAGGCTGTTGTCATAC; 5'-CAGGAAACAGCTATGACCATG), 10⁵ copies plasmid DNA, 0.8× Eva Green, and either 2 μl clarified lysate (neat or diluted 1:5) or 10–50 ng of purified His-tagged *Taq*. Primary hits displayed earlier C_{qs} compared to wild-type *Taq* controls processed in the same way on the same plates.

COLUMN PURIFICATION OF NON-TAGGED *Taq* MUTANTS

Mutant *pol* genes were subcloned into pET11 (with no His tag) and expressed in *E. coli* strain BL21-DE3-RIPL. One liter cultures were grown at 30°C in LB medium with 125 μg/ml ampicillin and 30 μg/ml chloramphenicol, and induced at OD_{600nm} of 0.6 with 1 mM IPTG for 4–5 h. Cell pellets were recovered by centrifugation and stored at –20°C. For purification, pellets were suspended in Buffer A (50 mM Tris-Cl pH 8.2, 1 mM EDTA, 10 mM 2-mercaptoethanol) plus cComplete Protease Inhibitor (Roche). Cell suspensions were disrupted by sonication, brought to 0.2 M with solid (NH₄)₂SO₄, heated in a water bath at 80°C for 15 min, and then cooled on ice. Polyethyleneimine was added to 0.2% (w/v), and after mixing thoroughly, insoluble material was removed by centrifugation. The supernatant was loaded on an SP Sepharose FastFlow (GE Healthcare) column equilibrated and run with Buffer B (50 mM Tris-Cl pH 8, 1 mM EDTA, 0.2 M (NH₄)₂SO₄, 10 mM 2-mercaptoethanol, 5% glycerol). Flow-through fractions were pooled and dialyzed against 15 volumes (with two changes) of Buffer C (50 mM Tris-Cl pH 8.3, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol), and then loaded on Q Sepharose HP (GE Healthcare) equilibrated with Buffer C. The column was eluted with a 12.5 column-volume gradient to Buffer C containing 400 mM KCl. *Taq*-containing fractions (as judged by SDS-PAGE) were pooled, diluted with 2.5 volumes of Buffer D (50 mM Tris-Cl pH 7.5, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, 125 mM KCl) and then loaded on Heparin Sepharose HP (GE Healthcare). The column was eluted with a 25 column-volume gradient to Buffer D with 650 mM KCl. Substantially-pure *Taq*-containing fractions

were pooled, dialyzed into storage buffer (20 mM Tris-Cl pH 8, 0.1 mM EDTA, 1 mM DTT, 100 mM KCl, 50% glycerol), and stored at -20°C . Protein was quantified using the Coomassie Plus protein assay (Thermo Fisher Scientific).

REAL-TIME PCR ASSAYS

SYBR Green qPCR reactions (20 μl) consisted of 10–20 ng wild-type or mutant *Taq*, 0.5–50 ng human genomic DNA, 200 μM each dNTP, 600 nM total primers, 1 \times *Taq* buffer (adjusted to 95 mM KCl for all PCRs and biochemical assays employing *Taq* mutants), and 0.24 \times SYBR Green. Primer sequences were as follows: ABC (F: 5'-CCAAACCCCTGGATCACGTGTT-3'; R: 5'-CCTCCCGCTCTCGTAGTTCT-3'), COMTE2 (F: 5'-GAGATCAACCCCCACTG-3'; R: 5'-GGCCCTTTTCCAG-3'), Quantos (F: 5'-TATAAGAAACTACTAACGCACCCAAAGG-3'; R: 5'-AAGAAAGGAGTCTAAGTGACTAACAG-3'; Aldolase (F: 5'-AGCCTAGCTCCAGTGCTTCTAGTA-3'; R: 5'-CTTGATGAGGAGCCGATATTG-3'), Numb (F: 5'-GAGGTTCCATA CAGGCACCTGCCAG-3'; R: 5'-CAAATCACCCCTCACAG TACTCTG-3'). TaqMan qPCR reactions consisted of 10 ng wild-type or mutant *Taq*, human genomic DNA, 0.8 mM dNTPs, 1 \times *Taq* buffer (adjusted to 95 mM KCl for *Taq* mutants), and 1 \times β -actin primer/probe (171 bp) from Life Technologies' Assays-On-Demand. qPCR reactions were run on the StepOnePlus (Life Technologies) or CFX96 (BioRad) instrument using cycling parameters indicated in the Figure legends.

PRIMER EXTENSION ASSAYS

Extension rate and processivity were measured at 70°C using M13mp18 template DNA (NEB), pre-annealed at a 1.3–1.5:1 molar ratio to a 5' ^{33}P -labeled primer with the sequence 5'-GGTTTCCCAGTCACGACGTTGAAACGACGGCCAGTGC-3'. Extension reactions (80 μl) consisting of 1.75 pmol primed M13, 200 μM each dNTP, and 1 \times *Taq* buffer were brought to 70°C prior to adding 0.55 pmol *Taq* enzyme. Aliquots (8 μl) were removed at 15, 30, 45, 60, 90, and 120 s and quenched with 50 mM EDTA. Extension products were denatured at 80°C for 3 min and analyzed on 1% alkaline-agarose gels to determine mean fragment length. Extension rates were calculated as number of nucleotides incorporated divided by incubation time. Processivity assays were conducted similarly using 1.1 pmol primed M13 and limiting amounts of *Taq* (template/enzyme molar ratios of 37:1, 370:1, 3700:1). Aliquots removed at various time points (15, 30, and 60 s) were quenched in gel loading buffer, and analyzed on 6% TBE-Urea gels (Life Technologies). Median processivity was determined from reactions producing the same product length over several time points or enzyme amounts.

NUCLEOTIDE INCORPORATION ASSAYS

Polymerase cocktails (10 μl) contained 200 μM dATP, dGTP and dCTP, 100 μM TTP (^{3}H -TTP), 1 \times *Taq* buffer (adjusted to 95 mM KCl for *Taq* mutants), and fixed or varying concentrations of primed (non-radiolabeled) M13 DNA for half-life ($T_{1/2}$) or K_{cat} measurements, respectively. Steady-state kinetic parameters were determined using 0.005 pmol *Taq* and 0.5–100 nM primed M13. After 3 min at 60°C , incorporation reactions were quenched with ice-cold 0.1 M EDTA and 5 μl aliquots spotted on DE-81 filters.

After washing 5 times with 2 \times SSC, incorporated radioactivity was measured by scintillation counting. K_m and V_{max} values were determined from Lineweaver-Burk plots, and K_{cat} was calculated as maximum number of nucleotides incorporated per *Taq* molecule per second ($V_{max}/[Taq]$). To determine $T_{1/2}(95^{\circ}\text{C})$, *Taq* was pre-heated in the absence or presence of genomic DNA before residual DNA polymerase activity was assayed as described above. Mixtures (10 μl) consisting of 0.02 pmol *Taq*, 1 \times *Taq* buffer, and 0 or 10 ng human genomic DNA were overlaid with mineral oil and incubated at 95°C for 5–180 min. At various time points, aliquots (2 μl) were transferred to polymerase assay cocktail (10 μl) containing 0.05 pmol primed M13. Polymerase activity (CPM) was plotted against pre-incubation time at 95°C .

K_d ASSAY

Dissociation constant K_d (DNA) was measured using a gel mobility-shift assay that employs a 5' ^{33}P -labeled hairpin template (5'-CTCCAGACACGACGAGTTGCCGATGGTC GACGTTCGCGAAAGCGAACGTCGACCACGGGCAACT-3') blocked at the 3' end with dideoxy TMP. The radiolabeled hairpin (0.25 nM) was incubated with varying concentrations of wild-type or mutant *Taq* DNA polymerases (0.5–1000 nM; bracketing the predicted K_d) in 1 \times *Taq* buffer at 37°C for 30 min. DNA-protein complexes were run on 10% non-denaturing TBE gels (Life Technologies) in 0.5 \times TBE buffer. Gels were dried down and exposed to film. The fraction of DNA bound was quantified by densitometry using AlphaView SA software (Alpha Innotech), and then plotted against enzyme concentration to determine K_d values by interpolation.

PCR/qPCR INHIBITION ASSAYS

Xylan, humic acid, CTAB, and dextran sulfate were purchased from Sigma. Blood was collected from healthy human volunteers using BD Vacutainer tubes with EDTA or heparin. Inhibitor-resistance screens ("spike-in" assays) were conducted by adding varying concentrations of inhibitors to qPCRs (20 μl) containing 20 ng wild-type or mutant *Taq* DNA polymerase, 200 μM each dNTP, 1 \times *Taq* buffer, 0.24 \times SYBR Green, 10^5 copies plasmid DNA, and 600 nM GAPDH primers (5'-ATCTTGAGGC TGTTGTCATAC; 5'-CAGGAAACAGCTATGACCATG) designed to amplify 549 bp GAPDH target on the Mx3005 qPCR system using 2 min initial denaturation at 95°C followed by 40 cycles of 12 s at 95°C and 60 s at 60°C .

Endpoint PCRs directly from blood (endogenous targets) consisted of (50 μl) 50 ng DNA polymerase, 15 mM Tris pH 8.8, 50 mM (wild-type *Taq*) or 95 mM (mutant *Taq*) KCl 2.5 mM MgCl₂, 0.02% Tween-20, 200 μM each dNTP, and 2% DMSO. A 322 bp IGF target was amplified directly from blood using 400 nM each primer (IGF322-F 5'-ATGGAGGGACCAATAGTAGGGAA; IGF322-R 5'-AGTACCACGTACAGGCTTGCAT), and the following cycling parameters: 5 min at 90°C followed by 30 cycles of 30 s at 95°C , 30 s at 60°C , 1 min at 72°C . Water (25 μl) was added to PCRs containing blood before centrifugation to pellet debris. A portion (12 μl) of each sample was run on a 4% Nusieve® (3:1) TBE agarose gel (Lonza). Comparisons employing commercial enzymes were conducted using the 232 bp Quantos

(see above) assay system and each manufacturer's recommended reaction buffer and cycling conditions.

RESULTS

Taq MUTAGENESIS AND SCREENING

We used the CSR technique to evolve mutants of *Taq* DNA polymerase that self-replicate using abbreviated extension times. Faster-replicating mutant polymerases are expected to provide robust performance with "fast" PCR instruments and cycling conditions. Moreover, identifying mutations that allow shorter cycle times may provide insight on kinetic factors that limit PCR performance of wild-type *Taq*.

Taq mutant libraries were subject to CSR selection using progressively shorter extension times, ranging from 1 min down to 6 s per kb of *taq pol I*. After rounds 2, 4, and 5, crude protein preparations were prepared and tested in amplifications employing 2-step cycling with 2 s denaturation and 7 s annealing-extension times. To account for differences in protein expression, lysates were prepared and assayed at least 2–3 times before affinity-purifying the top-performing *Taq* mutants. From a screen of several hundred clones, we recovered 24 His-tagged *Taq* mutants that consistently produce earlier C_q values compared to wild-type *Taq* in real-time PCR under fast cycling conditions. DNA sequencing identified 8 mutations that appear in 2 or more independent clones, as follows (in order of frequency): E507K (11); G59W (9); L245M (5); V155I, L375V, F749I (4); K508R, E734G (2).

To identify the most effective mutation or combination of mutations, we performed multi-site mutagenesis with an equimolar mixture of 8 mutant primers to create all possible (single, double, triple, etc.) combinations of G59W, V155I, L245M, L375V, E507K, K508R, E734G, and F749I mutations. The combinatorial library was enriched by 1 round of CSR selection using 15 s (6 s per kb) extension times, and random clones were screened as described above with a modified *Taq* PCR reaction buffer. We increased KCl concentration from 50 to 95 mM to eliminate non-specific amplification products that were generated by the majority of fast-cycling mutants (<30% of total amplified product as judged by agarose gels or qPCR melt curves; data not shown). Higher KCl concentrations (>50 mM) inhibit wild-type *Taq* DNA polymerase to the extent where no PCR products are generated at 95 mM KCl. The most active *Taq* mutants (1C2, 2C2, 3B, 42; see Table 1) were sub-cloned to remove the His-tag, purified using a standard wild-type *Taq* protocol, and characterized further by PCR.

In endpoint PCRs employing longer (>1 kb) genomic targets, *Taq* 1C2, 2C2, 3B, and 42 efficiently amplified targets using 15–30 s/kb extension times whereas *Taq* required 1 min/kb to

generate similar yields (data not shown). *Taq* 1C2, 2C2, 3B, and 42 were also tested with shorter targets to assess the benefits of fast cycling in quantitative PCR (qPCR) assays that traditionally employ targets of <300 bp (SYBR Green detection) or <200 bp (probe detection) to achieve amplification efficiencies as close to 100% as possible. In qPCRs employing SYBR Green, *Taq* readily amplifies genomic DNA targets up to 300 bp using standard anneal-extension times of 1 min at 60°C (Figure 1). However, with 10 s anneal-extension times, *Taq* produced slightly lower ΔR_n values in the 109 bp COMTE assay and completely failed to amplify a 305 bp NUMB target (Figures 2B,C). In contrast, *Taq* 1C2, 2C2, 3B, and 42 amplified the entire series (91–305 bp) of genomic DNA targets using abbreviated cycle times (10 s, Figure 2; 1 s, Figure 1), and no significant differences were observed among the mutants with respect to C_q s, ΔR_n values, and amplification efficiencies. Compared to SYBR Green assays, abbreviated cycle times appear to have less of an impact in TaqMan assays designed according to standard primer-probe design rules (<200 bp targets). In the example shown in Figure 3, *Taq* 1C2, 2C2, 3B, and 42 produce equivalent results to wild-type *Taq* using 10 s anneal-extension times, thereby confirming that probe-hydrolysis (5'-structure-specific endonuclease) activity is not affected by the fast-cycling mutations.

MUTATIONS CONFERRING THE FAST-CYCLING PHENOTYPE

DNA sequencing revealed that 3 of the original 8 mutations—L245M, E507K, and F749I—are absolutely conserved in *Taq* 1C2, 2C2, 3B, and 42, indicating that one or more are essential for fast-cycling under CSR selection conditions (Table 1). While not critical, other mutations emerging from CSR selection may further enhance overall fitness of the *Taq* 245M/507K/749I mutants. For example, G59W (missing in *Taq* 3B) and/or V155I (missing in *Taq* 42) appear in majority (3 out of 4) of the top-performers, suggesting a role for these N-terminal domain mutations in enhancing robustness. The L375V, K508R, and E734G mutations appear less frequently, and may be of limited importance to survival during CSR selection. Figure 4A shows the location of G59W, V155I, and L245M in the N-terminal 5'-3' endonuclease domain, and the close proximity of E507K and F749I in the thumb and fingers domains, respectively.

To further elucidate the contribution of individual mutations to the fast-cycling phenotype, we constructed and purified individual mutants *Taq* G59W, *Taq* V155I, *Taq* E507K, and *Taq* F749I. When equivalent amounts of enzyme were compared in qPCRs, only *Taq* E507K amplified 286 and 232 bp genomic DNA targets with short (1 s) anneal-extension times (Figure 1 and data not shown). Moreover, *Taq* E507K produced equivalent C_q values to

Table 1 | *Taq* mutations.

Enzyme	G59W	V155I	L245M	L375V	E507K	K508R	E734G	F749I
<i>Taq</i> (wt)	—	—	—	—	—	—	—	—
<i>Taq</i> 42	+	—	+	+	+	+	+	+
<i>Taq</i> 3B	—	+	+	—	+	—	—	+
<i>Taq</i> 2C2	+	+	+	+	+	—	+	+
<i>Taq</i> 1C2	+	+	+	—	+	—	—	+

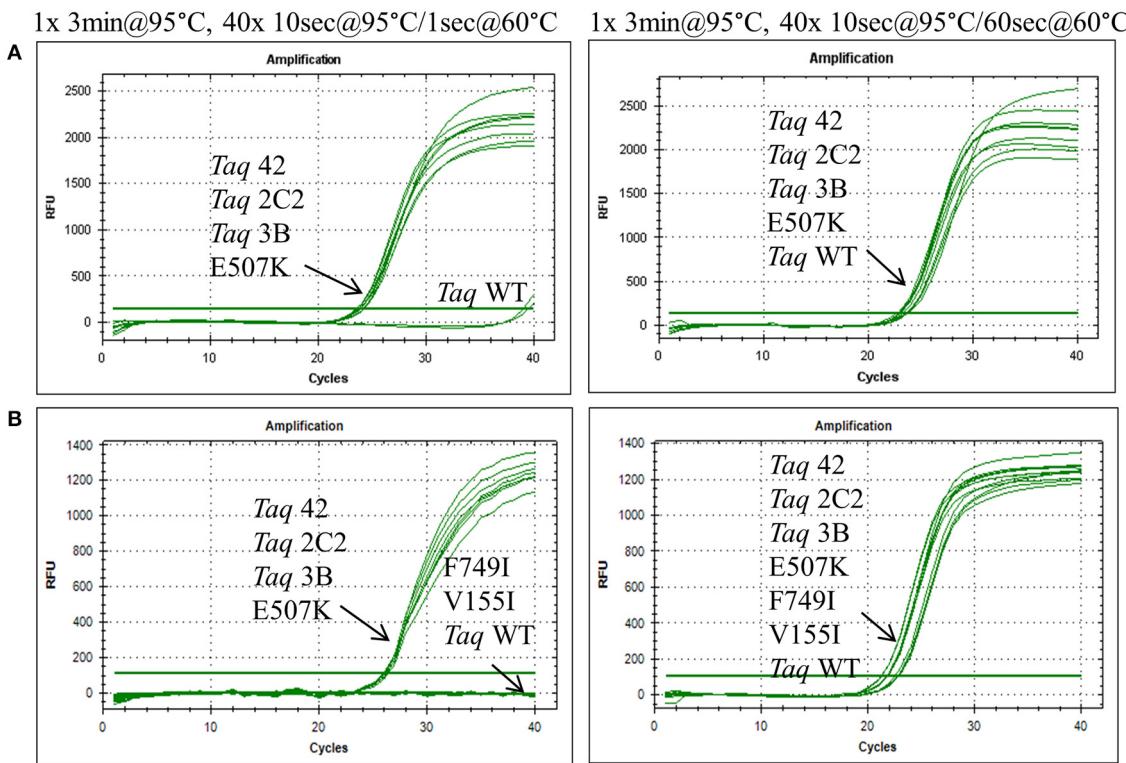


FIGURE 1 | Fast cycling conferred by E507K mutation. Genomic DNA targets were amplified as described in Methods using 10 ng human gDNA and 20 ng of purified wild-type or mutant *Taq* DNA polymerase. Reactions were cycled on the CFX96 instrument (BioRad)

using cycling conditions consisting of 3 min at 95°C followed by 40 cycles of (left panels) 10 s at 95°C, 1 s at 60°C; or (right panels) 10 s at 95°C, 60 s at 60°C. Genomic DNA targets are as follows: (A) 286 bp Aldolase; and (B) 232 bp Quantos.

Taq 42, 2C2, and 3B, indicating that the E507K mutation is solely responsible for the fast-cycling phenotype. With standard cycle times (60 s anneal-extension), amplification profiles are comparable across all *Taq* enzymes, confirming that single-site mutants *Taq* G59W, *Taq* V155I, and *Taq* F749I retain wild-type levels of polymerase activity.

KINETIC PARAMETERS OF *Taq* MUTANTS

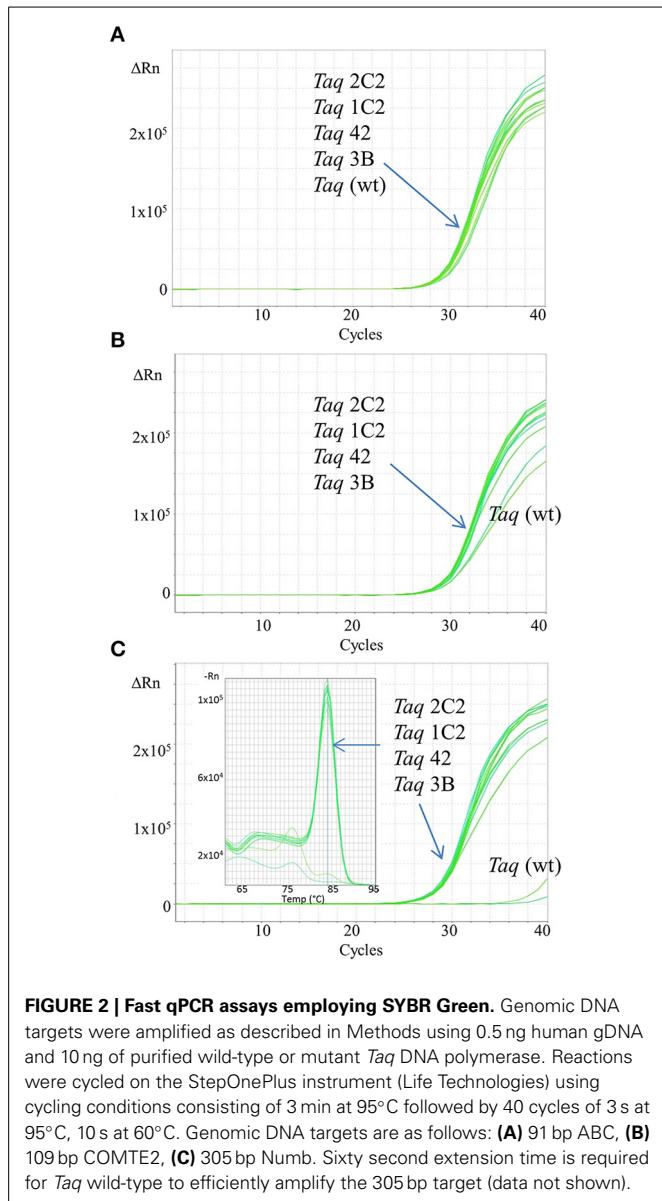
Kinetic parameters were investigated at each enzyme's KCl optimum (50 or 95 mM for *Taq* or *Taq* mutants, respectively) to determine how E507K confers the fast-cycling phenotype. In radio-labeled primer extension assays, *Taq* E507K and *Taq* 42 exhibit somewhat higher (1.7-fold) polymerization rates compared to wild-type *Taq* (85 vs. 50 nt s⁻¹; Table 2), but no change in processivity (20 bases; Table 2). Measurements of steady-state kinetic parameters also show a moderate (2.2-fold) increase in *K_{cat}* values (2.5 ± 0.03 s⁻¹ for *Taq* 42; 1.1 ± 0.04 s⁻¹ for *Taq*; data not shown). The most compelling difference between wild-type and E507K mutants was observed in gel-based *K_d* assays employing a hairpin oligonucleotide template. As shown in Table 2, dissociation constants for *Taq* 42, 1C2, and E507K mutants (2.75, 1.9, and 1.1 nM, respectively) were approximately 35–90-fold lower compared to those of *Taq* and *Taq* G59W (102 and 91 nM, respectively). Moreover, *K_d* measurements were comparable for assays run in the absence or presence of dNTPs (data not shown).

In sum, these results indicate that the E507K mutation supports faster PCR cycling conditions by increasing binding affinity of *Taq* for primed DNA templates irrespective of nucleotide binding.

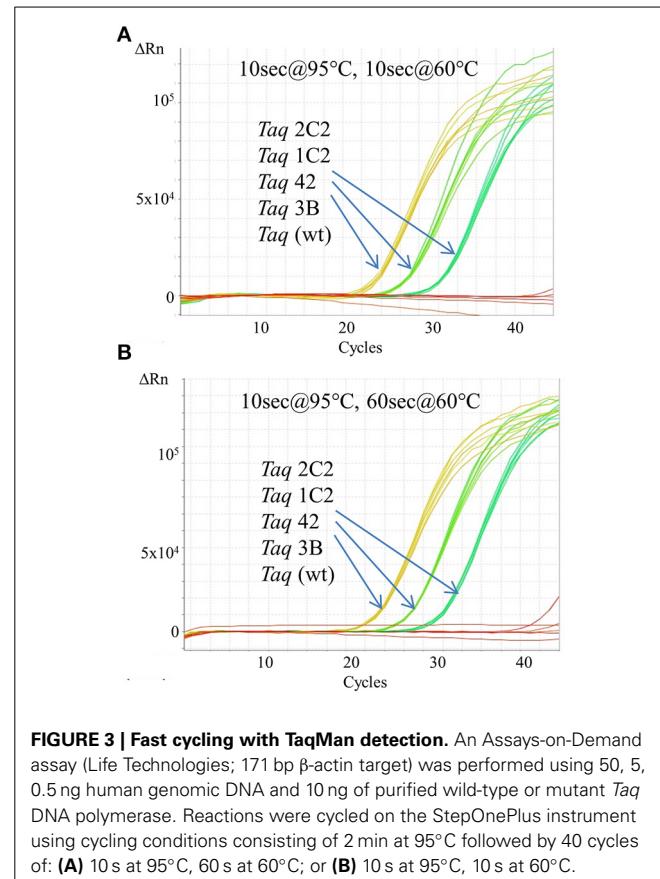
MUTATIONS IMPROVING POLYMERASE FITNESS

CSR has been shown to exert strong selective pressure on polymerase fitness, enriching for variants that self-replicate with sufficient accuracy and efficiency to remain in the gene pool (Ghadessy et al., 2001). In some cases, selected traits evolved in parallel with increased inhibitor resistance, presumably because survival requires self-replication in the presence of CSR emulsifiers/stabilizers and bacterial debris (Baar et al., 2011). In our study, several mutations including G59W, V155I, L245M, and F749I emerged during selection of the fast-cycling phenotype, prompting speculation that one or more of these ancillary mutations may enhance overall fitness of *Taq* DNA polymerase. To address this possibility, we further characterized the stability and inhibitor-resistance of a subset of *Taq* multi-site mutants.

Thermostability was investigated by determining half-life ($T_{1/2}$) at 95°C, in the absence and presence of DNA template. Compared to wild-type, *Taq* 42 exhibits a slightly higher $T_{1/2}$ in the presence of DNA (71 ± 4.6 vs. 60 ± 1.7 min), but no significant difference in the absence of DNA (60.5 ± 2.3 vs. 62 ± 2 min; data not shown). These results are consistent with increased



thermal resistance conferred by tighter binding (of E507K mutants) to DNA, but also imply that none of the other mutations in *Taq* 42 (G59W, V155I, L245M, L375V, K508R, E734G, F749I) enhance intrinsic thermal resistance. Next, inhibitor resistance of *Taq* 2C2 (all mutations except K508R) was investigated by amplifying an exogenous target in the presence of varying amounts of known *Taq* inhibitors, including plant-associated substances (xylan, dextran sulfate, CTAB), soil inhibitors (humic acid), whole blood treated with various anti-coagulants, and other PCR inhibitors (NaCl, SYBR Green) (Demeke and Adams, 1992; Watson and Blackwell, 2000; Kermekchiev et al., 2009). In these studies, inhibitor resistance was determined relative to wild-type *Taq* using standard PCR cycling times. Compared to wild-type *Taq*, *Taq* 2C2 shows significantly (≥ 4 -fold) higher tolerance to dextran sulfate (50-fold; data not shown), NaCl (4-fold; Table 3), and whole blood (see below).



Additional testing with single- and multi-site mutants revealed that E507K confers increased tolerance to NaCl (up to 100 mM) in addition to higher affinity for primed-template. In contrast, mutations at other positions result in equivalent (V155I, F749) or reduced (*Taq* G59W; by 2.5-fold) tolerance to NaCl compared to wild-type *Taq*. Next, resistance to blood-associated inhibitors was investigated in more depth by amplifying endogenous targets directly from blood (EDTA tubes). As shown in Figure 5, the E507K mutation also provides increased tolerance to blood, allowing amplification of a 322 bp IGF target from up to 45% EDTA-blood. Resistance to blood may be partly explained by higher salt tolerance, as the amount of NaCl introduced with 22.5 μ l blood is within the range (up to 100 mM; Table 3) tolerated by *Taq* E507 (67.5 mM NaCl final). Interestingly, individual mutations at G59W, V155I, and F749I also confer significant resistance to EDTA-blood, albeit less than E507K (15% compared to 45%), perhaps reflecting intrinsic sensitivity to NaCl (tolerate <10–25 mM, while 15% blood introduces 22.5 mM NaCl). In contrast, wild-type *Taq* fails to amplify the 322 bp endogenous target from as little as 1% EDTA-blood (Table 3), consistent with previous reports (Al-Soud and Radstrom, 1998; Kermekchiev et al., 2009). In total, these results indicate that G59W, V155I, and F749I mutations confer significant (>10–15-fold) resistance to blood-associated inhibitors through a mechanism that is distinct from E507K (unrelated to increased binding affinity and salt resistance). Moreover, when the majority of CSR-selected mutations

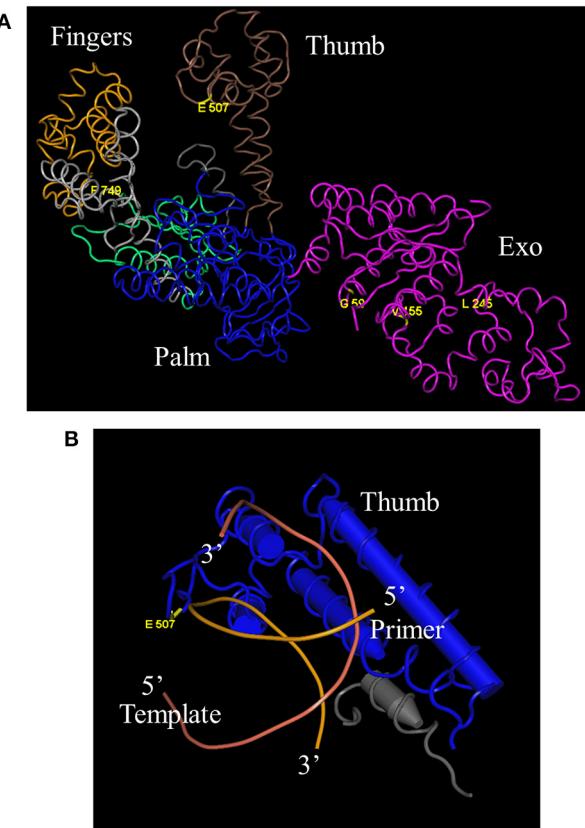


FIGURE 4 | Location of E507K and blood-resistant mutations in *Taq* crystal structure. “Fitness mutations” G59W, V155I, L245M, and F749I reside in the 5'-3' exonuclease and Fingers domains, respectively, while fast-cycling mutation E507K is in the thumb domain (**A**) (Kim et al., 1995). E507 resides in close proximity to the primer portion of the primer-template duplex (**B**) (Li et al., 1998).

Table 2 | Kinetic parameters of wild-type and mutant *Taq* polymerases.

Enzyme	Extension rate (nt/s)	Processivity (nt)	K_d [DNA] (nM)
<i>Taq</i> (wt)	50	20	102
<i>Taq</i> 42	85	20	2.75
<i>Taq</i> 1C2	ND	ND	1.9
E507K	85	20	1.1
G59W	ND	ND	91.6

ND, Not Determined.

are added to E507K, resistance increases from 45% (*Taq* E507K) to between 50% (*Taq* 3B; 3 additional mutations) and 60–65% (*Taq* 1C2, 2C2, 42; 4–6 additional mutations).

The fast-cycling *Taq* mutants were found to be much less resistant to heparinized blood. The difference was more pronounced for single-site mutants ($\leq 2\%$ for heparinized blood vs. 15–45% EDTA-blood) than for combinatorial mutants (10–30% vs. 50–65% for heparin- vs. EDTA-blood, respectively). Presumably, heparinized blood poses a greater challenge due

Table 3 | Resistance of *Taq* mutants to blood and NaCl*.

Enzyme	EDTA blood (%)	Heparinized blood (%)	NaCl (mM)
<i>Taq</i> (wt)	<1	<1	25
<i>Taq</i> 42	60	10	100
<i>Taq</i> 3B	50	10	100
<i>Taq</i> 2C2	65	25	100
<i>Taq</i> 1C2	60	30	100
E507K	45	2	100
F749I	15	2	25
V155I	15	<2	25
G59W	15	2	10

*The values represent the highest inhibitor concentration beyond which PCR amplification is inhibited.

to the combined effects of these known inhibitors (blood and heparin; Satsangi et al., 1994). Resistance to heparinized blood decreases in the following order: 1C2, 2C2 (25–30%) > 42, 3B (10%) > G59W, E507K, F749I (2%) > V155I ($\geq 1\%$) > wild-type *Taq* (<1%; Table 3), indicating that additional mutations (along with E507K) are essential to overcoming the inhibitory burden posed by heparinized blood. *Taq* 1C2 DNA polymerase has been incorporated into a commercially-available master mix (SureDirect PCR, Agilent Technologies), designed for amplification of genomic targets including cfDNA from large volumes of blood. When compared to other blood PCR kits, we found the majority can amplify the 232 bp Quantos target from 2.5 and 25% EDTA- and heparin-treated blood (Figure 6). Only the *Taq* 1C2-based formulation could amplify an endogenous target from 45% blood (heparin-treated).

DISCUSSION

In this report, we describe the properties of *Taq* DNA polymerase mutants evolved to self-replicate under abbreviated cycle times. Mutations emerging from CSR selection (G59W, V155I, L245M, L375V, E507K, K508R, E734G, F749I) were randomly combined and subjected to a final round of selection employing 15 s anneal-extension times. The fastest-cycling combinatorial mutants (*Taq* 42, 1C2, 2C2, and 3B) were shown to readily amplify genomic DNA targets up to 300 bp using abbreviated (≤ 10 s hold times) two-step cycling protocols on fast-ramping instruments (Figures 1, 2). When tested individually, the fast-cycling phenotype could be attributed solely to the E507K mutation (Figure 1B). Further characterization revealed that compared to wild-type, *Taq*E507K exhibits a dramatic (~ 90 -fold) reduction in K_d for primer-template (measured in the absence of nucleotides), and only moderate (~ 1.7 -fold) or no improvement in polymerization rate or processivity, respectively. These findings suggest that *Taq*’s affinity for primed-template, rather than catalytic rate, is a limiting factor in PCR amplification under fast-cycling conditions.

A previous study concluded that E507 plays a role in primer-template binding, as substitution of E for Q improves *Taq*’s RNA-dependent 5' nuclease activity without altering DNA-dependent 5' nuclease activity or RNA- and DNA-dependent DNA polymerase activities (Ma et al., 2000). This result would

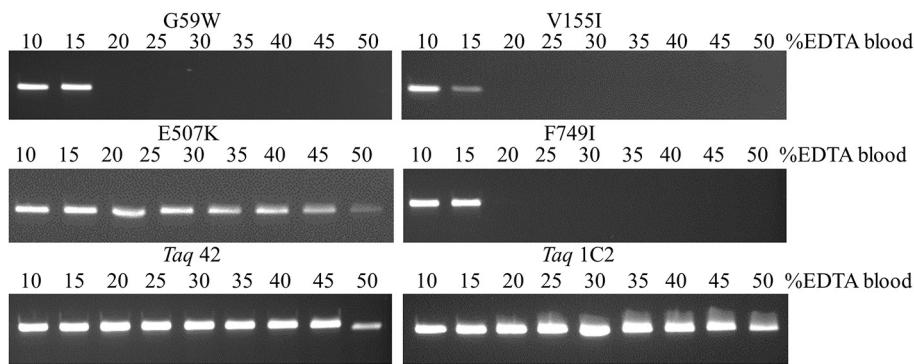


FIGURE 5 | Amplification directly from blood. A 322 bp IGF target was amplified from 5 to 25 μ l EDTA-blood using 50 ng of each mutant *Taq* DNA polymerase. Reactions were cycled on the SureCycler 8800 using the following parameters: 5 min at 90°C, followed by 30 cycles of 30 s at 95°C,

30 s at 60°C, 60 s at 72°C. Wild-type *Taq* can amplify the 322 bp from human genomic DNA in the absence of blood, but not in the presence of 1% blood (data not shown), even though the amount of EDTA introduced with 0.5 μ l EDTA-blood (0.089 mM EDTA) is well below inhibitory levels.

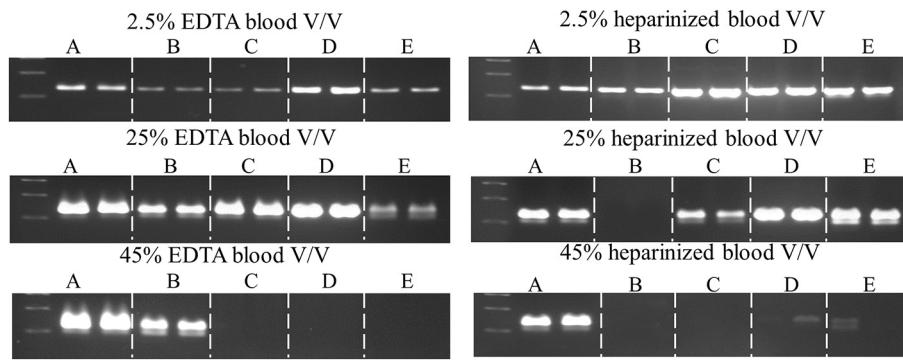


FIGURE 6 | Comparison of commercial enzymes for blood PCR. A 232 bp single-copy target was amplified (50 μ l) from the indicated volumes of blood using SureDirect PCR mix (**A**) or another commercially-available PCR blood kits (**B–E**). All reactions contained 0.4 μ M each Quantos primer and were cycled on a SureCycler 8800 instrument, according to each manufacturer's

recommendations. PCRs performed with SureDirect PCR master mix were cycled as follows: 5 min at 90°C, followed by 30 cycles of 30 s at 94°C, 30 s at 60°C, 60 s at 72°C. The other commercial kits employ 3-step cycling and 30 cycles, with hold times ranging between 5 and 60 s (annealing) and 15–60 s (extension), respectively.

be expected if the amino acid side chain at position 507 comes in close proximity to the template strand, and substituting amine for negatively-charged oxygen reduces discrimination against the 2'OH of the ribose sugar. The crystal structure of *Taq* large fragment shows that E507 resides in the H1H2 loop of the thumb domain, which interacts with the distal portion of the primer-template duplex in both open and closed forms of binary and ternary complexes (Li et al., 1998). In this model, the peptide carbonyl of E507 comes in close proximity (3.8 Å) to the phosphate moiety between the 6th and 7th nucleotide from the 3' end of the newly-extended primer (Figure 4B). As a whole, these data suggest that the E507K mutation stabilizes the *Taq*-DNA binary complex by forming additional contacts with the distal (away from the active site) portion of the primed-template. Higher KCl (95 mM instead of 50 mM) may be required with *Taq* E507K mutants to reduce binding affinity for mis-annealed primer-template during PCR annealing steps. The importance of E507 residue in protein/primer-template interactions has been shown in other DNA polymerase families as well. For example, in a study by Cozens et al. (2012), an analogs mutation to *Taq*

E507K was made in the thumb domain of *Thermococcus gorgonarius* (family B DNA polymerase; E664K), which transforms *Tgo* into an RNA polymerase by lowering the K_d for non-cognate RNA/DNA duplex and lowering the K_m for ribonucleotide incorporation. This mutant was also capable of translesion synthesis across an abasic site or thymidine dimer.

As discussed above, CSR has been shown to exert strong selective pressure on polymerase fitness, enriching for variants that self-replicate with sufficient fidelity and catalytic efficiency to remain in the gene pool. For example, selected traits such as improved thermostability and increased tolerance to inhibitors have been shown to evolve with no cost to fidelity or catalytic efficiency (K_{cat}/K_m) (Ghadessy et al., 2001; Baar et al., 2011). CSR selection for tolerance to individual inhibitors has also been shown to produce broad spectrum resistance, presumably because survival requires self-replication in the presence of CSR emulsifiers, stabilizers, and bacterial debris (e.g., denatured protein, nucleic acid, and membrane lipid in the aqueous PCR compartments). In a striking example, CSR selection with bone powder produced a chimeric polymerase (2D9) with broad

tolerance to a variety of environmental inhibitors, including humic acid, coprolite, peat extract, clay-rich soil, cave sediment, and tar, but surprisingly not to inhibitors in whole blood (Baar et al., 2011). Despite 81 amino acid changes, 2D9 exhibited comparable fidelity and processivity to wild-type *Taq*, consistent with the intrinsic requirement for polymerase fitness. A broad resistance spectrum implies a common mechanism of inhibition (for bone and soil extracts), prompting the authors to speculate that non-specific binding of inhibitors to protein and/or nucleic acid may sequester *Taq* or DNA template and prevent PCR amplification (Baar et al., 2011).

Encouraged by these reports and others, we assayed our CSR-selected mutants for inhibitor resistance. In addition to fast-cycling, the E507K mutation was shown to improve resistance to NaCl (tolerates up to 100 mM) and to inhibitors in whole blood (tolerates up to 45% (v/v) EDTA-blood). Individual mutations G59W, V155I, and F749I also confer blood resistance, but the magnitude of improvement is less (by 3-fold) than for E507K, and no corresponding increase in NaCl tolerance was observed (*Taq* L245M not tested). Heparinized blood posed a significant challenge due to the combined inhibitory effects of blood and heparin, as shown by the drastic difference in blood volumes tolerated by *Taq* E507K (22.5 μ l EDTA-blood vs. 1 μ l heparin-blood per 50 μ l PCR). Apparently, increased DNA-binding affinity and NaCl-resistance conferred by the E507K substitution is insufficient to overcome the inhibitory burden, and additional mutations are required for amplification from larger ($\geq 5 \mu$ l) volumes of heparinized blood. Not all mutations or mutation combinations have been tested, but our results to date implicate G59W and 1–3 additional mutations (V155I, L245M, and/or F749I) in improving tolerance to heparinized blood. Compared to *Taq* E507K, *Taq* 3B (E507K plus V155I, L245M, F749I) can amplify genomic targets directly from 5% or 8% more EDTA- or heparin-blood, respectively. Adding the G59W mutation (*Taq* 3B plus G59W) further improves tolerance, allowing *Taq* 1C2 to amplify from 15% (EDTA-treated) or 28% (heparin-treated) more blood compared to *Taq* E507K.

Heparin and hemoglobin/hemin (the most potent inhibitors in blood) are thought to mimic and compete with duplex DNA for binding to the polymerase active site (Byrnes et al., 1975; Akane et al., 1994; Satsangi et al., 1994; Ghadessy et al., 2001). *Taq* mutants with increased tolerance toward heparin or blood have been described previously. After mutagenizing four amino acids implicated in cold-sensitivity and overall performance, Kermekchiev et al. identified several substitutions at E708 that dramatically (>30-fold) improve resistance of *Taq* E626K/I707L and KlenTaq (N-truncated *Taq*) E626K/I707L to blood, hemoglobin/hemin, soil extract, and humic acid (Kermekchiev et al., 2009). Although no kinetic data were provided, results of a competitive filter-binding assay suggested that generic resistance to PCR inhibitors was related to increased affinity for DNA rather than diminished binding to hemin or humic acid (Kermekchiev et al., 2009). With the possible exception of humic acid, *Taq* E507K mutant exhibits a similarly-broad spectrum of resistance compared to wild-type (*Taq* E507K vs. *Taq* E626K/I707L/E708Q: >45- vs. ~100-fold for EDTA-blood; 0 vs. 2-fold for SYBR Green; 10- vs. 32-fold for SYBR in the presence

of blood; 2- vs. 8-fold for humic acid) in addition to increased affinity for primer-template. E708 mutants have been commercialized by DNA Polymerase Technology, Inc. under the trade names OmniTaq (*Taq* E626K/I707L/E708Q) and OmniKlenTaq (KlenTaq E626K/I707L/E708K) (Zhang et al., 2010).

As discussed above, E507 resides in the primer-template binding site where mutations are expected to modulate DNA binding affinity. It's less clear how substitutions at or near E708 contribute to cold-sensitivity or increased DNA affinity/inhibitor resistance, as this region lies at the hinge-point of the fingers, away from the fingertip (which contacts incoming nucleotide and single-stranded template) and portions of the thumb and palm domains that interact with primer-template (Kermekchiev et al., 2009). *Taq* E507K and E708Q mutants are further distinguished by relative resistance to heparinized blood. *Taq* E507K tolerates up to 45% EDTA-blood compared to 2% heparin-blood, suggesting the combination of heparin and blood-associated (e.g., hemoglobin/hemin) inhibitors saturates a common (mutually-exclusive) DNA/heme/heparin binding site. Alternatively, E507K confers blood resistance by lowering sensitivity to NaCl, and mutations at other positions confer resistance to DNA mimics by reducing binding affinity for heme/heparin. In contrast to *Taq* E507K, E708 mutants (OmniTaq, OmniKlenTaq) exhibit similar tolerances to blood treated with EDTA, heparin, and citrate, consistent with a unique mechanism of inhibitor resistance (Zhang et al., 2010).

The heparin-binding site of *Taq* was precisely delineated by Ghadessy et al. (2001), who used CSR selection with heparin to identify a *Taq* mutant (*Taq* H15) with 130-fold higher resistance to heparin and comparable (to wild-type *Taq*) affinity for DNA in BIACore and template-dilution assays. Heparin resistance-conferring mutations cluster in the DNA-binding domain, and four of six residues mutated in *Taq* H15 make direct contacts with either the primer (K540, N583) or template (D578, M747) strand in open and closed forms of the binary/ternary complex (Ghadessy et al., 2001).

By selecting for increased speed rather than blood/heparin resistance, we and Kermekchiev et al. (2009) identified mutations that confer increased DNA binding affinity and tolerance for blood; curiously, the E507K mutant retained sensitivity to heparin, while E708 mutants appear equally tolerant to EDTA-blood and heparinized blood. In contrast, Ghadessy et al. (2001) identified 6 mutations in H15 that collectively diminish heparin binding without altering K_d (DNA), even though heparin is a DNA mimic and the DNA-heparin binding sites are thought to overlap. These findings can be reconciled if one assumes that *Taq* makes additional contacts with primer-template outside the DNA/heparin binding site. In the Ghadessy study, one or more of the H15 mutations (K540, D578, N583, M747) may lower affinity for DNA and heparin through the same mechanism, while the remaining mutations restore K_d (equal to wild-type *Taq*) through additional interactions with primer-template. Residing close to K540, N583 and the 3' primer terminus, E507 is close to the DNA-heparin binding site, and introducing a positively-charged side chain (E→K) may increase binding affinity for both DNA and heparin (lower heparin tolerance). In contrast to E507, E708 lies farther from primer-template binding site, where mutations that

increase DNA binding affinity won't necessarily enhance heparin binding or sensitivity.

Among the other blood-resistant mutations identified here, F749I is of particular interest due to its location relative to the template strand. Although pointing inward, F749 is flanked by amino acids that interact directly with heparin (M747) or the template nucleoside opposing the incoming nucleoside triphosphate (R746, M747, N750). F749 is also thought to stack against I707 (Kermekchiev et al., 2003), which may provide a causal link to the cold-sensitive (KlenTaq 706–708 mutants) and inhibitor-resistant (*Taq*/KlenTaq 708 mutants) phenotypes identified by Kermekchiev et al. (2003, 2009).

The contribution of mutations located outside the DNA-binding pocket is more difficult to rationalize. In this study, individual mutations G59W and V155I confer increased (by >15-fold) resistance to EDTA-blood, while G59W improves amplification from heparinized blood (by 3.5-fold) when added to *Taq* V155I/L245M/E507K/F749I. G59 and V155 reside in the 5' structure-specific nuclease domain responsible for excising Okazaki RNA during lagging strand synthesis (Li et al., 1998). The 5' nuclease domain was implicated in an earlier study showing that N-truncated *Taq* (KlenTaq; residues 279–882) can amplify from 10- to 100-fold more blood (up to 5–10%) compared to full-length *Taq* (Kermekchiev et al., 2009). KlenTaq also exhibits higher thermostability and lower processivity compared to wild-type, suggesting that point mutations in the 5' nuclease domain can modulate several properties of *Taq* that contribute to PCR performance (Barnes, 1992). Mutations that increase PCR fitness may alleviate the effects of other blood-associated inhibitors (immunoglobulin G, lactoferrin, proteases) that inhibit PCR by unknown means (Al-Soud and Radstrom, 1998, 2001; Al-Soud et al., 2000).

REFERENCES

- Akane, A., Matsubara, K., Nakamura, H., Takahashi, S., and Kimura, K. (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J. Forensic Sci.* 39, 362–372.
- Al-Soud, W. A., Jonsson, L. J., and Radstrom, P. (2000). Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR. *J. Clin. Microbiol.* 38, 345–350.
- Al-Soud, W., and Radstrom, P. (1998). Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Appl. Environ. Microbiol.* 64, 3748–3753.
- Al-Soud, W. A., and Radstrom, P. (2001). Purification and characterization of PCR-inhibitory components in blood cells. *J. Clin. Microbiol.* 39, 485–493. doi: 10.1128/JCM.39.2.485–493.2001
- Baar, C., d'Abbadie, M., Vaisman, A., Arana, M. E., Hofreiter, M., Woodgate, R., et al. (2011). Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Res.* 39, e51. doi: 10.1093/nar/gkq1360
- Barnes, W. M. (1992). The fidelity of *Taq* polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* 112, 29–35. doi: 10.1016/0378-1119(92)90299-5
- Byrnes, J. J., Downey, K. M., Esserman, L., and So, A. G. (1975). Mechanism of hemin inhibition of erythroid cytoplasmic DNA polymerase. *Biochemistry* 14, 796–799. doi: 10.1021/bi00675a023
- Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R., and Holliger, P. (2012). A short adaptive path from DNA to RNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8067–8072. doi: 10.1073/pnas.1120964109
- d'Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., et al. (2007). Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* 25, 939–943. doi: 10.1038/nbt1321
- Demeke, T., and Adams, R. P. (1992). The effects of plant polysaccharides and buffer additives on PCR. *Biotechniques* 12, 332–334.
- Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4552–4557. doi: 10.1073/pnas.071052198
- Hashimoto, M., Chen, P. C., Mitchell, M. W., Nikitopoulos, D. E., Soper, S. A., and Murphy, M. C. (2004). Rapid PCR in a continuous flow device. *Lab Chip* 4, 638–645. doi: 10.1039/b406860b
- Hilscher, C., Vahrson, W., and Dittmer, D. P. (2005). Faster quantitative real-time PCR protocols may lose sensitivity and show increased variability. *Nucleic Acids Res.* 33, e182. doi: 10.1093/nar/gni181
- Holland, P. M., Abramson, R. D., Watson, R., and Gelfand, D. H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 88, 7276–7280. doi: 10.1073/pnas.88.16.7276
- Kermekchiev, M. B., Kirilova, L. I., Vail, E. E., and Barnes, W. M. (2009). Mutants of *Taq* DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Res.* 37, e40. doi: 10.1093/nar/gkn1055
- Kermekchiev, M. B., Tzakov, A., and Barnes, W. M. (2003). Cold-sensitive mutants of *Taq* DNA polymerase provide a hot start for PCR. *Nucleic Acids Res.* 31, 6139–6147. doi: 10.1093/nar/gkg813
- Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* 376, 612–616. doi: 10.1038/376612a0
- Li, Y., Korolev, S., and Waksman, G. (1998). 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.* 17, 7514–7525. doi: 10.1093/emboj/17.24.7514
- Ma, W. P., Kaiser, M. W., Lyamicheva, N., Schaefer, J. J., Allawi, H. T., Takova, T., et al. (2000). RNA template-dependent 5' nuclease activity of *Thermus aquaticus* and *Thermus thermophilus* DNA polymerases. *J. Biol. Chem.* 275, 24693–24700. doi: 10.1074/jbc.M002268200
- Satsangi, J., Jewell, D. P., Welsh, K., Bunce, M., and Bell, J. I. (1994). Effect of heparin on polymerase chain reaction. *Lancet* 343, 1509–1510. doi: 10.1016/S0140-6736(94)92622-0
- Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* 32, 1197–1207. doi: 10.1093/nar/gkh271
- Watson, R. J., and Blackwell, B. (2000). Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Can. J. Microbiol.* 46, 633–642. doi: 10.1139/w00-043
- Zhang, Z., Kermekchiev, M. B., and Barnes, W. M. (2010). Direct DNA amplification from crude clinical samples using a PCR enhancer cocktail and novel mutants of *Taq*. *J. Mol. Diagn.* 12, 152–161. doi: 10.2353/jmoldx.2010.090070

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 May 2014; paper pending published: 02 July 2014; accepted: 18 July 2014; published online: 14 August 2014.

Citation: Arezi B, McKinney N, Hansen C, Cayouette M, Fox J, Chen K, Lapira J, Hamilton S and Hogrefe H (2014) Compartmentalized self-replication under fast PCR cycling conditions yields *Taq* DNA polymerase mutants with increased DNA-binding affinity and blood resistance. *Front. Microbiol.* 5:408. doi: 10.3389/fmicb.2014.00408

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Arezi, McKinney, Hansen, Cayouette, Fox, Chen, Lapira, Hamilton and Hogrefe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Bacteriophage T7 DNA polymerase – sequenase

Bin Zhu*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Kirk Matthew Schnorr, Novozymes A/S, Denmark

Samir Hamdan, King Abdullah University of Science and Technology, Saudi Arabia

***Correspondence:**

Bin Zhu, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, C2-221, 240 Longwood Avenue, Boston, MA 02115, USA
e-mail: bin_zhu@hms.harvard.edu

An ideal DNA polymerase for chain-terminating DNA sequencing should possess the following features: (1) incorporate dideoxy- and other modified nucleotides at an efficiency similar to that of the cognate deoxynucleotides; (2) high processivity; (3) high fidelity in the absence of proofreading/exonuclease activity; and (4) production of clear and uniform signals for detection. The DNA polymerase encoded by bacteriophage T7 is naturally endowed with or can be engineered to have all these characteristics. The chemically or genetically modified enzyme (Sequenase) expedited significantly the development of DNA sequencing technology. This article reviews the history of studies on T7 DNA polymerase with emphasis on the serial key steps leading to its use in DNA sequencing. Lessons from the study and development of T7 DNA polymerase have and will continue to enlighten the characterization of novel DNA polymerases from newly discovered microbes and their modification for use in biotechnology.

Keywords: bacteriophage T7, DNA polymerase, sequenase, DNA sequencing, marine phages

INITIAL CHARACTERIZATION

DNA polymerases catalyze the synthesis of DNA, a pivot process in both living organisms and in biotechnology (Hamilton et al., 2001; Reha-Krantz, 2008). Family A DNA polymerases including *Escherichia coli* DNA polymerase I, Taq DNA polymerase, and T7 DNA polymerase have served as prototypes for biochemical and structural studies on DNA polymerases and have been widely used as molecular reagents (Patel et al., 2001; Loh and Loeb, 2005).

A DNA polymerase activity from bacteriophage T7 was first observed in an *E. coli* mutant deficient in DNA polymerase I infected with bacteriophage T7 (Grippo and Richardson, 1971). The initial characterization of T7 DNA polymerase was intriguing. Although the gene responsible for the polymerase activity was mapped to gene 5 (Hinkle and Richardson, 1974; Hori et al., 1979b), gene 5 protein (gp5) itself had what appeared to be no DNA polymerase activity but only ssDNA exonuclease activity (Hori et al., 1979a). Apparently a host component was required to reconstitute the full DNA polymerase (Modrich and Richardson, 1975a). This host factor turned out to be a small redox protein – *E. coli* thioredoxin (Modrich and Richardson, 1975b; Mark and Richardson, 1976). The redox capacity of thioredoxin, however, is not required for stimulation of the DNA polymerase activity (Huber et al., 1986). Instead thioredoxin plays a structural role in stabilizing the binding of gene 5 protein to a primer-template (Huber et al., 1987) and increase the processivity of the polymerase more than 100-fold (Tabor et al., 1987a), representing a unique function of this universal protein. Thioredoxin binds to a 71-residue loop of T7 gene 5 protein (Doublie et al., 1998), which is not present in other Pol I-type polymerases, resulting in a stable 1:1 complex ($K_D = 5 \text{ nM}$; Tabor et al., 1987a).

Another intriguing finding during the initial characterization of T7 DNA polymerase is on its exonuclease activity. T7 DNA polymerase lacks the 5'-3' exonuclease activity found in *E. coli* DNA polymerase I but does possess a strong 3'-5' single and

double stranded DNA exonuclease activity (Hori et al., 1979b). The double-stranded DNA exonuclease activity requires the presence of thioredoxin. Interestingly, various protein purification procedures, depending on the presence or absence of EDTA in the buffer, can generate T7 DNA polymerases that differ significantly in their exonuclease activity, resulting in two forms of DNA polymerase (Fischer and Hinkle, 1980; Engler et al., 1983). By comparison of the two forms of polymerase and careful tracking of the purification procedures, it was revealed that the exonuclease activity of T7 DNA polymerase could be specifically inactivated in an oxidation reaction by oxygen, a reducing agent and ferrous ion (Tabor and Richardson, 1987b). The easily modifiable exonuclease and extraordinary processivity of T7 DNA polymerase kindled the emergence of a powerful tool in the DNA sequencing era.

SEQUENASE ERA

Invented by Sanger et al. (1977), the method of chain-terminating sequencing initiated a revolution toward the genome-sequencing era. However, the enzymes initially used for chain-terminating sequencing, the Klenow fragment of *E. coli* DNA polymerase I and avian myeloblastosis virus (AMV) reverse transcriptase, had low processivity (~15 nt for Klenow fragment and 200 for AMV reverse transcriptase, the latter has a relatively higher processivity but its rate of DNA synthesis is only several nucleotides per second). Processivity describes the number of nucleotides continuously incorporated by a DNA polymerase using the same primer-template without dissociation. Thus if the DNA polymerase used for chain-terminating sequencing is non-processive, artifactual bands will arise at positions corresponding to the nucleotide at which the polymerase dissociated. Frequent dissociation will create strong background that obscures the true DNA sequence. Although the issue can be partially improved by long time incubation with high concentration of substrates that may “chase” those artifactual bands up to higher molecular weight,

this procedure is not an ideal solution since reinitiation of primer elongation at dissociation sites (usually regions of compact secondary structure or hairpins) is inefficient and may result in the incorporation of incorrect nucleotides. Although T7 DNA polymerase itself has a processivity of only a few nucleotides, the association with *E. coli* thioredoxin dramatically increases its processivity. Consequently, with T7 DNA polymerase termination of a sequencing reaction will occur only at positions where a chain-terminating agent (such as a dideoxynucleotide) is incorporated, yielding a long DNA sequence (Tabor and Richardson, 1987c).

A more severe problem with DNA polymerases used prior to T7 DNA polymerase is the discrimination against dideoxynucleotides, the chain-terminating nucleotides used in Sanger sequencing. Most of known DNA polymerases strongly discriminate against ddNTP. For example, T4 DNA polymerase, *E. coli* DNA polymerase I, Taq DNA polymerase, and Vent DNA polymerase incorporate a dideoxynucleoside monophosphate (ddNMP) at least a 1000 times slower than the corresponding deoxynucleoside monophosphate (dNMP). To use these polymerases in DNA sequencing a high ratio of ddNTP to dNTP must be used for efficient chain-termination. Even though the overall incorporation of ddNMP can be improved in such an uneconomic way, wide variation in the intensity of adjacent fragments still occur because the extent of discrimination varies with different DNA sequences and structures. T7 DNA polymerase, however, is at the other end of the spectrum, discriminating against ddNTP only several-fold. Thus a much lower concentration of ddNTP can be used with T7 DNA polymerase and the uniformity of DNA bands on the gel is much higher (Tabor and Richardson, 1987c). The discrimination was further lowered by replacing magnesium with manganese in the sequencing reaction (Tabor and Richardson, 1989a). With Mn^{2+} in an isocitrate buffer, T7 DNA polymerase incorporates dNMP and ddNMP at same rate, resulting in uniform terminations of sequencing reactions.

With the naturally endowed high processivity and the lack of discrimination against ddNTP, the only hindrance for T7 DNA polymerase as a DNA sequencing enzyme is its robust 3'-5' exonuclease activity. Exonuclease activity increases the fidelity of DNA synthesis by excising newly synthesized bases incorrectly base-paired to the template. For applications like PCR it is often a desired feature. While for DNA sequencing such activity is detrimental since when the dNTP concentration falls, the rate of exonuclease activity increases close to that of polymerase activity, resulting in no net DNA synthesis or degradation of DNA. The associated exonuclease activity will also cause DNA polymerase to idle at regions with secondary structures in the template, producing variability in the intensity of signals. The iron-catalyzed oxidation mentioned above can produce modified T7 DNA polymerase with greatly reduced exonuclease activity, and this chemically modified enzyme was the basis for Sequenase and the first easy-to-use DNA sequencing kits commercialized by United States Biochemical Co. However, the residual exonuclease activity can still result in some loss of labeled DNA bands upon prolonged incubation (Tabor and Richardson, 1987b). Tabor and Richardson carried out an extensive chemical and mutagenesis screen for selective elimination of the

exonuclease activity of T7 DNA polymerase. The rapid screen of a large number of mutants was based on the observation that exonuclease minus mutants of T7 DNA polymerase can synthesize through a specific hairpin region in the DNA template (Tabor and Richardson, 1989b). As a result many mutants deficient in exonuclease activity were revealed and among them a mutant lacking 28 amino acids in the N-terminal exonuclease domain had no detectable exonuclease activity, while its polymerase activity is significantly higher that of the wild-type protein. This mutant was the basis of version 2 of Sequenase. Sequenase pioneered development of thermostable enzymes and facilitated the automation for high-throughput sequencing.

Degradation of a DNA fragment can occur via a nucleophilic attack on the 3'-terminal internucleotide linkage by H_2O or pyrophosphate (PPi). The 3'-5' exonuclease catalyzes the former reaction, generating dNMP or ddNMP. The latter reaction is called pyrophosphorolysis. As the reversal of polymerization, pyrophosphorolysis generates dNTP or ddNTP, sometimes resulting in "holes": the disappearance of ddNMP labeled DNA fragments on the gel. By adding pyrophosphatase to the reaction to cleave PPi the pyrophosphorylation can be eliminated (Tabor and Richardson, 1990). The combination of modified T7 DNA polymerase, manganese ion, and pyrophosphatase can generate accurate and uniform bands on a DNA sequencing gel to the extent that, the DNA sequence can be directly determined by the relative intensity of each band if different amount of the four ddNTPs are added at certain ratio (Tabor and Richardson, 1990).

Thermostability is a highly desired feature for DNA polymerase. A thermostable enzyme like Taq DNA polymerase is superior for cycle sequencing, in which multiple rounds of DNA synthesis are carried out from the same template, with the newly synthesized DNA strand released after each cycle by heat denaturation. The heat stable DNA polymerase survives the denaturation step and is available for the next cycle of polymerization. Cycle sequencing allows much less DNA template and polymerase to be used in a sequencing reaction. In cycle sequencing low processivity is an advantage because a polymerase with low processivity cycles rapidly, decreasing the chance of strong specific stops. However, the strong discrimination against ddNTP (at least 100-fold, often 10,000-fold) by most thermostable DNA polymerase was a significant obstacle for their use in cycle sequencing. Although the use of manganese ion can decrease the discrimination (Tabor and Richardson, 1989a), manganese has several disadvantages compared with magnesium such as narrow working concentration, precipitation, and less activity of DNA polymerase than that supported by magnesium ion.

Studies on T7 DNA polymerase led to one of the most elegant demonstrations of enzyme engineering and turned Taq DNA polymerase into "Thermo Sequenase." To pursue the molecular mechanism underlying the discrepancy in discrimination against ddNTP among family A DNA polymerases, Tabor and Richardson swapped the five most conserved regions in the crevice responsible for binding DNA and NTPs between T7 DNA polymerase and *E. coli* DNA polymerase I (Tabor and Richardson, 1995), based on the 3D structure of *E. coli* DNA polymerase I. By an SDS-DNA activity assay, the "Helix O" from *E. coli* DNA polymerase

I was observed to confer strong discrimination against ddNTP to T7 DNA polymerase. Further mutagenesis in this region revealed that the tyrosine-526 in T7 DNA polymerase or the homologous position phenylalanine-762 in *E. coli* DNA polymerase I was the single determinant for discrimination against ddNTP. When the corresponding residue, F667 in Taq DNA polymerase was replaced with tyrosine, the modified Taq DNA polymerase F667Y actually preferred ddNTP 2-fold over dNTP, comparing to the 6000-fold discrimination against ddNTP by the wide-type enzyme (Tabor and Richardson, 1995). Taq DNA polymerase F667Y, with its naturally endowed superior thermostability and engineered elimination of discrimination against ddNTP, was the basis for “Thermo Sequenase,” an enzyme that greatly expedited the Human Genome Project. The structure of T7 DNA polymerase in complex with a primed-template and a nucleoside triphosphate solved later (Doublie et al., 1998) revealed that the 3'-hydroxyl of the incoming nucleotide and the hydroxyl of Tyr 526 are both within hydrogen-bonding distance of the pro S_p-oxygen of the β-phosphate and suggested that one or both of these interactions may be required for nucleotide incorporation. However, even with the structure one could not have predicted the dramatic effect of tyrosine-526 on nucleotide analog discrimination.

AN IDEAL MODEL TO STUDY INTERACTIONS WITHIN A REPLICOSOME

T7 DNA polymerase consisting of T7 gene 5 protein and *E. coli* thioredoxin, together with T7 gp4 bifunctional primase-helicase, and gene 2.5 ssDNA-binding proteins constitute the simplest known replisome that mediates coordinated leading- and lagging-strand DNA synthesis (Richardson, 1983; Debyser et al., 1994; Lee et al., 1998; Hamdan and Richardson, 2009). The concise organization of the T7 replisome makes it ideal for studies of the multiple interactions of DNA polymerase during the movement of the replisome such as loading of the polymerase (Zhang et al., 2011), polymerase exchange (Johnson et al., 2007), processive synthesis (Hamdan et al., 2007), and translesion synthesis (Zhu et al., 2011). Critical interactions for coordinated DNA synthesis including polymerase-thioredoxin (Johnson and Richardson, 2003; Ghosh et al., 2008; Akabayov et al., 2010; Tran et al., 2012), polymerase-helicase (Zhang et al., 2011; Kulczyk et al., 2012), polymerase-primase (Chowdhury et al., 2000; Zhu et al., 2010), and polymerase-gene 2.5 single-stranded DNA binding protein (He et al., 2003; Hamdan et al., 2005; Ghosh et al., 2009, 2010) interaction were extensively studied. The solid biochemical background of T7 DNA polymerase also attracted investigations using single-molecular methods (Lee et al., 2006; Hamdan et al., 2009; Pandey et al., 2009; Etson et al., 2010; Loparo et al., 2011; Geertsema et al., 2014).

NOVEL T7-LIKE DNA POLYMERASES

DNA polymerases from microbes advanced DNA sequencing technology that in turn unveiled a much larger, diverse and unexplored microbial world. Metagenomics data indicates that the marine phages are the most abundant and diverse organisms on the earth (Suttle, 2005), of which 60–80% potential gene products do not match any in the database. A large portion

of these gene products must be involved in the nucleic acid metabolism, thus one can expect numerous novel nucleic acid enzymes that can enrich the present toolbox of enzymes derived from a small group of characterized microbes. Indeed, our own initial effort on the characterization of marine phage polymerases have revealed unique features of a single-subunit RNA polymerase from marine cyanophage Syn5 that can complement the predominantly used T7 RNA polymerase for *in vitro* RNA synthesis (Zhu et al., 2013a,b). Characterization of marine phage DNA polymerases appears more promising since one can easily target numerous interesting DNA polymerases from the reported marine phage genomes, even just for T7-like or family A DNA polymerases such as those from cyanophage Syn5 (Pope et al., 2007) and P-SSP7 (Sullivan et al., 2005), phages infecting SAR116-clade bacterium (Kang et al., 2013) and marine ssDNA phages (Schmidt et al., 2014). Considering the high probability that the 60–80% unmatched genes may harbor novel polymerase genes, the marine phage is an unlimited treasure to contribute new polymerase tools that can fulfill niches in biotech industry. Characterization and engineering of T7 DNA polymerase has shown the value of identifying novel properties of nucleic acid enzymes.

REFERENCES

- Akabayov, B., Akabayov, S. R., Lee, S. J., Tabor, S., Kulczyk, A. W., and Richardson, C. C. (2010). Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, *Escherichia coli* thioredoxin. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15033–15038. doi: 10.1073/pnas.1010141107
- Chowdhury, K., Tabor, S., and Richardson, C. C. (2000). A unique loop in the DNA-binding crevice of bacteriophage T7 DNA polymerase influences primer utilization. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12469–12474. doi: 10.1073/pnas.230448397
- Debyser, Z., Tabor, S., and Richardson, C. C. (1994). Coordination of leading and lagging strand DNA synthesis at the replication fork of bacteriophage T7. *Cell* 77, 157–166. doi: 10.1016/0092-8674(94)90243-7
- Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391, 251–258. doi: 10.1038/34593
- Engler, M. J., Lechner, R. L., and Richardson, C. C. (1983). Two forms of the DNA polymerase of bacteriophage T7. *J. Biol. Chem.* 258, 11165–11173.
- Etson, C. M., Hamdan, S. M., Richardson, C. C., and van Oijen, A. M. (2010). Thioredoxin suppresses microscopic hopping of T7 DNA polymerase on duplex DNA. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1900–1905. doi: 10.1073/pnas.0912664107
- Fischer, H., and Hinkle, D. C. (1980). Bacteriophage T7 DNA replication in vitro. Stimulation of DNA synthesis by T7 RNA polymerase. *J. Biol. Chem.* 255, 7956–7964.
- Geertsema, H. J., Kulczyk, A. W., Richardson, C. C., and van Oijen, A. M. (2014). Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. *Proc. Natl. Acad. Sci. U.S.A.* doi: 10.1073/pnas.1402010111 [Epub ahead of print].
- Ghosh, S., Hamdan, S. M., Cook, T. E., and Richardson, C. C. (2008). Interactions of *Escherichia coli* thioredoxin, the processivity factor, with bacteriophage T7 DNA polymerase and helicase. *J. Biol. Chem.* 283, 32077–32084. doi: 10.1074/jbc.M805062200
- Ghosh, S., Hamdan, S. M., and Richardson, C. C. (2010). Two modes of interaction of the single-stranded DNA-binding protein of bacteriophage T7 with the DNA polymerase-thioredoxin complex. *J. Biol. Chem.* 285, 18103–18112. doi: 10.1074/jbc.M110.107656
- Ghosh, S., Marintcheva, B., Takahashi, M., and Richardson, C. C. (2009). C-terminal phenylalanine of bacteriophage T7 single-stranded DNA-binding protein is essential for strand displacement synthesis by T7 DNA polymerase at a nick in DNA. *J. Biol. Chem.* 284, 30339–30349. doi: 10.1074/jbc.M109.024059

- Grippo, P., and Richardson, C. C. (1971). Deoxyribonucleic acid polymerase of bacteriophage T7. *J. Biol. Chem.* 246, 6867–6873.
- Hamdan, S. M., Johnson, D. E., Tanner, N. A., Lee, J. B., Qimron, U., Tabor, S., et al. (2007). Dynamic DNA helicase-DNA polymerase interactions assure processive replication fork movement. *Mol. Cell* 27, 539–549. doi: 10.1016/j.molcel.2007.06.020
- Hamdan, S. M., Loparo, J. J., Takahashi, M., Richardson, C. C., and van Oijen, A. M. (2009). Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* 457, 336–339. doi: 10.1038/nature07512
- Hamdan, S. M., Marintcheva, B., Cook, T., Lee, S. J., Tabor, S., and Richardson, C. C. (2005). A unique loop in T7 DNA polymerase mediates the binding of helicase-prime, DNA binding protein, and processivity factor. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5096–5101. doi: 10.1073/pnas.0501637102
- Hamdan, S. M., and Richardson, C. C. (2009). Motors, switches, and contacts in the replisome. *Annu. Rev. Biochem.* 78, 205–243. doi: 10.1146/annurev.biochem.78.072407.103248
- Hamilton, S. C., Farchaus, J. W., and Davis, M. C. (2001). DNA polymerases as engines for biotechnology. *Biotechniques* 31, 370–376.
- He, Z. G., Rezende, L. F., Willcox, S., Griffith, J. D., and Richardson, C. C. (2003). The carboxyl-terminal domain of bacteriophage T7 single-stranded DNA-binding protein modulates DNA binding and interaction with T7 DNA polymerase. *J. Biol. Chem.* 278, 29538–29545. doi: 10.1074/jbc.M304318200
- Hinkle, D. C., and Richardson, C. C. (1974). Bacteriophage T7 deoxyribonucleic acid replication in vitro. Requirements for deoxyribonucleic acid synthesis and characterization of the product. *J. Biol. Chem.* 249, 2974–2980.
- Hori, K., Mark, D. F., and Richardson, C. C. (1979a). Deoxyribonucleic acid polymerase of bacteriophage T7. Characterization of the exonuclease activities of the gene 5 protein and the reconstituted polymerase. *J. Biol. Chem.* 254, 11598–11604.
- Hori, K., Mark, D. F., and Richardson, C. C. (1979b). Deoxyribonucleic acid polymerase of bacteriophage T7. Purification and properties of the phage-encoded subunit, the gene 5 protein. *J. Biol. Chem.* 254, 11591–11597.
- Huber, H. E., Russel, M., Model, P., and Richardson, C. C. (1986). Interaction of mutant thioredoxins of *Escherichia coli* with the gene 5 protein of phage T7. The redox capacity of thioredoxin is not required for stimulation of DNA polymerase activity. *J. Biol. Chem.* 261, 15006–15012.
- Huber, H. E., Tabor, S., and Richardson, C. C. (1987). *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. *J. Biol. Chem.* 262, 16224–16232.
- Johnson, D. E., and Richardson, C. C. (2003). A covalent linkage between the gene 5 DNA polymerase of bacteriophage T7 and *Escherichia coli* thioredoxin, the processivity factor: fate of thioredoxin during DNA synthesis. *J. Biol. Chem.* 278, 23762–23772. doi: 10.1074/jbc.M301366200
- Johnson, D. E., Takahashi, M., Hamdan, S. M., Lee, S. J., and Richardson, C. C. (2007). Exchange of DNA polymerases at the replication fork of bacteriophage T7. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5312–5317. doi: 10.1073/pnas.0701062104
- Kang, I., Oh, H. M., Kang, D., and Cho, J. C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12343–12348. doi: 10.1073/pnas.1219930110
- Kulczyk, A. W., Akabayov, B., Lee, S. J., Bostina, M., Berkowitz, S. A., and Richardson, C. C. (2012). An interaction between DNA polymerase and helicase is essential for the high processivity of the bacteriophage T7 replisome. *J. Biol. Chem.* 287, 39050–39060. doi: 10.1074/jbc.M112.410647
- Lee, J., Chastain, P. D. 2nd, Kusakabe, T., Griffith, J. D., and Richardson, C. C. (1998). Coordinated leading and lagging strand DNA synthesis on a minicircular template. *Mol. Cell* 1, 1001–1010. doi: 10.1016/S1097-2765(00)80100-8
- Lee, J. B., Hite, R. K., Hamdan, S. M., Xie, X. S., Richardson, C. C., and van Oijen, A. M. (2006). DNA primase acts as a molecular brake in DNA replication. *Nature* 439, 621–624. doi: 10.1038/nature04317
- Loh, E., and Loeb, L. A. (2005). Mutability of DNA polymerase I: implications for the creation of mutant DNA polymerases. *DNA Repair (Amst.)* 4, 1390–1398. doi: 10.1016/j.dnarep.2005.09.006
- Loparo, J. J., Kulczyk, A. W., Richardson, C. C., and van Oijen, A. M. (2011). Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3584–3589. doi: 10.1073/pnas.1018824108
- Mark, D. F., and Richardson, C. C. (1976). *Escherichia coli* thioredoxin: a subunit of bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 73, 780–784. doi: 10.1073/pnas.73.3.780
- Modrich, P., and Richardson, C. C. (1975a). Bacteriophage T7 Deoxyribonucleic acid replication in vitro. A protein of *Escherichia coli* required for bacteriophage T7 DNA polymerase activity. *J. Biol. Chem.* 250, 5508–5514.
- Modrich, P., and Richardson, C. C. (1975b). Bacteriophage T7 deoxyribonucleic acid replication in vitro. Bacteriophage T7 DNA polymerase: an enzyme composed of phage- and host-specific subunits. *J. Biol. Chem.* 250, 5515–5522.
- Pandey, M., Syed, S., Donmez, I., Patel, G., Ha, T., and Patel, S. S. (2009). Coordinating DNA replication by means of priming loop and differential synthesis rate. *Nature* 462, 940–943. doi: 10.1038/nature08611
- Patel, P. H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L. A. (2001). Prokaryotic DNA polymerase I: evolution, structure, and “base flipping” mechanism for nucleotide selection. *J. Mol. Biol.* 308, 823–837. doi: 10.1006/jmbi.2001.4619
- Pope, W. H., Weigle, P. R., Chang, J., Pedulla, M. L., Ford, M. E., Houtz, J. M., et al. (2007). Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a “horned” bacteriophage of marine *Synechococcus*. *J. Mol. Biol.* 368, 966–981. doi: 10.1016/j.jmb.2007.02.046
- Reha-Krantz, L. J. (2008). Recent patents of gene sequences relative to DNA polymerases. *Recent Pat. DNA Gene Seq.* 2, 145–163. doi: 10.2174/187221508786241729
- Richardson, C. C. (1983). Bacteriophage T7: minimal requirements for the replication of a duplex DNA molecule. *Cell* 33, 315–317. doi: 10.1016/0092-8674(83)90411-7
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467.
- Schmidt, H. F., Sakowski, E. G., Williamson, S. J., Polson, S. W., and Wommack, K. E. (2014). Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. *ISME J.* 8, 103–114. doi: 10.1038/ismej.2013.124
- Sullivan, M. B., Coleman, M. L., Weigle, P., Rohwer, F., and Chisholm, S. W. (2005). Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* 3:e144. doi: 10.1371/journal.pbio.0030144
- Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160
- Tabor, S., and Richardson, C. C. (1989a). Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *J. Biol. Chem.* 264, 6447–6458.
- Tabor, S., and Richardson, C. C. (1989b). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci. U.S.A.* 86, 4076–4080. doi: 10.1073/pnas.86.11.4076
- Tabor, S., and Richardson, C. C. (1990). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Effect of pyrophosphorylation and metal ions. *J. Biol. Chem.* 265, 8322–8328.
- Tabor, S., and Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6339–6343. doi: 10.1073/pnas.92.14.6339
- Tabor, S., Huber, H. E., and Richardson, C. C. (1987a). *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* 262, 16212–16223.
- Tabor, S., and Richardson, C. C. (1987b). Selective oxidation of the exonuclease domain of bacteriophage T7 DNA polymerase. *J. Biol. Chem.* 262, 15330–15333.
- Tabor, S., and Richardson, C. C. (1987c). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 84, 4767–4771. doi: 10.1073/pnas.84.14.4767
- Tran, N. Q., Lee, S. J., Akabayov, B., Johnson, D. E., and Richardson, C. C. (2012). Thioredoxin, the processivity factor, sequesters an exposed cysteine in the thumb domain of bacteriophage T7 DNA polymerase. *J. Biol. Chem.* 287, 39732–39741. doi: 10.1074/jbc.M112.409235

- Zhang, H., Lee, S. J., Zhu, B., Tran, N. Q., Tabor, S., and Richardson, C. C. (2011). Helicase-DNA polymerase interaction is critical to initiate leading-strand DNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9372–9377. doi: 10.1073/pnas.1106678108
- Zhu, B., Lee, S. J., and Richardson, C. C. (2010). Direct role for the RNA polymerase domain of T7 primase in primer delivery. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9099–9104. doi: 10.1073/pnas.1004220107
- Zhu, B., Lee, S. J., and Richardson, C. C. (2011). Bypass of a nick by the replisome of bacteriophage T7. *J. Biol. Chem.* 286, 28488–28497. doi: 10.1074/jbc.M111.252023
- Zhu, B., Tabor, S., Raytcheva, D. A., Hernandez, A., King, J. A., and Richardson, C. C. (2013a). The RNA polymerase of marine cyanophage Syn5. *J. Biol. Chem.* 288, 3545–3552. doi: 10.1074/jbc.M112.442350
- Zhu, B., Tabor, S., and Richardson, C. C. (2013b). Syn5 RNA polymerase synthesizes precise run-off RNA products. *Nucleic Acids Res.* 42:e33. doi: 10.1093/nar/gkt1193

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 March 2014; paper pending published: 31 March 2014; accepted: 01 April 2014; published online: 16 April 2014.

Citation: Zhu B (2014) Bacteriophage T7 DNA polymerase – sequenase. *Front. Microbiol.* 5:181. doi: 10.3389/fmicb.2014.00181

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Engineering processive DNA polymerases with maximum benefit at minimum cost

Linda J. Reha-Krantz^{1*}, Sandra Woodgate² and Myron F. Goodman³

¹ Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

² Trevigen, Inc., Gaithersburg, MD, USA

³ University of Southern California, Los Angeles, CA, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Sylvie Doublet, University of Vermont, USA

Michael Trakselis, Baylor University, USA

*Correspondence:

Linda J. Reha-Krantz, Department of Biological Sciences, University of Alberta, CW405 Biosciences Building, Edmonton, AB T6G 2E9, Canada
e-mail: linda.reha-krantz@ualberta.ca

DNA polymerases need to be engineered to achieve optimal performance for biotechnological applications, which often require high fidelity replication when using modified nucleotides and when replicating difficult DNA sequences. These tasks are achieved for the bacteriophage T4 DNA polymerase by replacing leucine with methionine in the highly conserved Motif A sequence (L412M). The costs are minimal. Although base substitution errors increase moderately, accuracy is maintained for templates with mono- and dinucleotide repeats while replication efficiency is enhanced. The L412M substitution increases intrinsic processivity and addition of phage T4 clamp and single-stranded DNA binding proteins further enhance the ability of the phage T4 L412M-DNA polymerase to replicate all types of difficult DNA sequences. Increased pyrophosphorolysis is a drawback of increased processivity, but pyrophosphorolysis is curbed by adding an inorganic pyrophosphatase or divalent metal cations, Mn²⁺ or Ca²⁺. In the absence of pyrophosphorolysis inhibitors, the T4 L412M-DNA polymerase catalyzed sequence-dependent pyrophosphorolysis under DNA sequencing conditions. The sequence specificity of the pyrophosphorolysis reaction provides insights into how the T4 DNA polymerase switches between nucleotide incorporation, pyrophosphorolysis and proofreading pathways. The L-to-M substitution was also tested in the yeast DNA polymerases delta and alpha. Because the mutant DNA polymerases displayed similar characteristics, we propose that amino acid substitutions in Motif A have the potential to increase processivity and to enhance performance in biotechnological applications. An underlying theme in this chapter is the use of genetic methods to identify mutant DNA polymerases with potential for use in current and future biotechnological applications.

Keywords: A+T- and G+C-rich DNA templates, DNA polymerase-DNA dynamics, DNA polymerase processivity, DNA polymerase translocation, DNA sequencing, motif A, pyrophosphorolysis, replication fidelity

INTRODUCTION

The remarkable advances made in molecular biology in the last 40 years are dependent to a large extent on DNA polymerase-dependent methods. Researchers who pioneered the use of DNA polymerases for DNA sequencing, PCR and site-directed mutagenesis were awarded Nobel prizes for their ground-breaking achievements, but DNA polymerases play key roles in numerous additional indispensable methods including DNA labeling, cloning, whole genome amplification, and diagnostic techniques. Researchers, however, recognized from the earliest days that some DNA polymerase activities are counter-productive for *in vitro* applications. Nuclease activities, for example, can degrade primers and the newly synthesized DNA, yet some exonuclease activity is needed for high fidelity DNA replication. Other DNA polymerase activities need to be modified, for example the ability to utilize non-standard nucleotides and to replicate “difficult” DNA templates with simple repeats or sequences that are excessively rich in A+T or G+C base pairs. DNA polymerases from several organisms have been subject to extensive engineering in order to remove or curb unwanted

activities and to modify or enhance others that are required for optimal performance *in vitro* (see several chapters in this volume and reviews by Hamilton et al., 2001; Reha-Krantz, 2008).

One of the most challenging tasks has been to engineer DNA polymerases to replicate difficult DNA sequences. DNA polymerases stall and often dissociate at difficult template sequences and these sequences are frequently sites for replication errors. For example (Streisinger et al., 1966; Streisinger and Owen, 1985) observed that simple repeat sequences in phage T4 genes are hotspots *in vivo* for frameshift mutations, which are typically insertions or deletions within mono- and di-nucleotide repeat sequences. The length of the repeat sequence affects the likelihood of frameshift mutations as hotspot mutation sites have longer repeat tracks than colder, less mutable sites. Streisinger et al. (1966) and Streisinger and Owen (1985) that frameshift mutations are created by transient separation of the primer and template strands followed by strand misalignment during reannealing to create an intermediate with an unpaired repeat in the template or primer

strand, but with a correctly paired primer-template terminus that can be extended by a DNA polymerase. If the misaligned DNA strands are not corrected before the next round of chromosome replication, the repeat sequence will be expanded (if the unpaired repeat is in the primer strand) or contracted (if the unpaired repeat is in the template strand). As expected, these “slippery” DNA templates are also difficult to replicate *in vitro* as they slow or even prevent further replication and are sites for insertions and deletions (Rao, 1994; Fidalgo da Silva and Reha-Krantz, 2000; Clarke et al., 2001; Fazekas et al., 2010). Kunkel et al. (1994) demonstrated that +1 insertions in homopolymeric runs dramatically increase in reactions with the phage T7 DNA polymerase in the absence of thioredoxin, the T7 DNA polymerase processivity factor. Thus, decreased DNA polymerase processivity is correlated with increased strand misalignment mutagenesis, presumably by increasing DNA polymerase dissociation which provides an opportunity for the free primer-end to spontaneously denature and then to re-anneal in a misaligned configuration. This proposal is supported by other studies, for example, Fazekas et al. (2010).

We observed that the leucine to methionine (L412M) substitution in the conserved Motif A (**Table 1**) of the bacteriophage T4 DNA polymerase produced a mutant DNA polymerase that has improved ability to replicate difficult DNA sequences under typical DNA sequencing conditions. Furthermore, the L412M-DNA polymerase has increased ability to incorporate and extend modified nucleotides. Both activities are facilitated by increased intrinsic processivity (Reha-Krantz and Nonay, 1994), as explained below.

The L412M substitution in the phage T4 DNA polymerase was identified by a genetic selection strategy as a second-site amino acid substitution that suppressed the excessive proofreading observed for several mutant DNA polymerases (Stocki et al., 1995; Li et al., 2010). As expected, the L412M substitution reduces proofreading activity, but base substitution mutations are increased moderately 10- to 40-fold, while frameshift mutations by only threefold at most (**Table 2**). Thus, high fidelity DNA replication observed for the wild type T4 DNA polymerase, about one error in 10^7 – 10^8 nucleotides incorporated (Kunkel et al., 1984), is essentially retained; hence, the T4 L412M-DNA polymerase is an excellent candidate polymerase for single-molecule sequencing and other applications where error-free replication is required. Note that amino acid substitutions in Motif A of DNA polymerases from several organisms have been observed to cause minor to major changes in replication fidelity, both increasing and decreasing accuracy (**Table 2**; Reha-Krantz and Nonay, 1994; Patel and Loeb, 2000; Patel et al., 2001; Fidalgo da Silva et al., 2002; Niimi et al., 2004; Li et al., 2005; Venkatesan et al., 2006, 2007; Pursell et al., 2007; Sakamoto et al., 2007; Nick McElhinny et al., 2008; Zhong et al., 2008).

We review the development of the phage T4 L412M-DNA polymerase as an important DNA sequencing tool and its use in other applications. There are useful lessons to be learned because replication of difficult DNA templates remains salient with current DNA sequencing and amplification technologies.

Because amino acid substitutions in Motif A of the bacteriophage T4 DNA polymerase have profound effects on intrinsic processivity (**Figure 1**), we propose that amino acid substitutions in Motif A of other DNA polymerases, especially the L-to-M substitution in family B DNA polymerases, have the potential to increase processivity and enhance DNA polymerase performance in biotechnological applications. A strategy to “evolve” mutant DNA polymerases with increased processivity and other desirable activities for *in vitro* applications (Bourn et al., 2011) is compared to our genetic evolution strategies. We also present new data about pyrophosphorylation activity, a byproduct of increased intrinsic processivity. While the scope of this chapter is limited to methods used to enhance the intrinsic processivity of the phage T4 DNA polymerase, information about the naturally highly processive ϕ 29 DNA polymerase is presented by Blanco et al. (1983) and Eid et al. (2009). The association of thioredoxin with the T7 DNA polymerase and the development of Sequenase™ is presented by Zhu (2014). Increased processivity produced by fusing DNA-binding domains to DNA polymerases is discussed by Arezi et al. (2014) in this volume.

MATERIALS AND METHODS

DNA POLYMERASES AND DNA POLYMERASE ACCESSORY PROTEINS

Expression, purification, and characterization of wild type and mutant T4 DNA polymerases were described previously (Lin et al., 1987; Spicer et al., 1988; Reha-Krantz et al., 1993). The T4 gp44 clamp and clamp loading proteins (gp44/gp62) and the T4 single-stranded DNA (ssDNA) binding protein (gp32) were expressed and purified as described (Shamoo et al., 1986; Rush et al., 1989), but with modifications for large scale production.

DNA POLYMERASE INTRINSIC PROCESSIVITY

The intrinsic processivity of wild type and mutant phage T4 DNA polymerases was determined in the presence of a heparin trap as described by Reha-Krantz and Nonay (1994). Briefly, 20 μ l reactions contained 25 mM HEPES, pH 7.5, 60 mM NaOAc, 1 mM dithiothreitol, 0.5 mM EDTA, 80 μ M dNTPs, 0.2 mg/ml bovine serum albumin, 7.5 nM primed single-stranded circular M13 DNA (expressed as 3'-primer ends), and 150 nM DNA polymerase. Reactions were pre-incubated 5 min at 30°C and started by adding a solution of Mg²⁺ [Cf, 6 mM Mg(OAc)₂] and heparin (Cf, 0.1 mg/ml). Reactions were incubated 15 s at 30°C and stopped by the addition of 2 μ l 0.2 M EDTA. Reaction products were separated on DNA sequencing gels (7% acrylamide, 8 M urea) and the ³²P-labeled products were visualized by exposure to Kodak X-Omat AR film.

PYROPHOSPHORYLATION ASSAY CONDITIONS AND INHIBITION BY Mn²⁺ AND Ca²⁺ IONS

Pyrophosphorylation activity was measured for the exonuclease-deficient D112A/E114A/L412M-DNA polymerase in 12 μ l reactions containing partially digested duplex DNA that was 3'-labeled with [³²P]dCMP (50 pmol of labeled 3'-ends/reaction), 10 nM DNA polymerase, 67 mM Tris-HCl (pH 7.5), 16.7 mM (NH₄)₂SO₄, 0.5 mM DTT, 167 μ g BSA/ml, 6.7 mM MgCl₂, and

Table 1 | Motif A in Phage T4 DNA polymerase and yeast DNA polymerases.

DNA pol	Motif A sequence	Phenotypes
[15pt]T4 wild type	408 D It SLYPSII	Wild type
T4 L412M	M	Weak mutator ^a , strongly PAA-sensitive ^a
T4 I417V	V	Antimutator ^a , sensitive to reduced dGTP ^a
T4 L412M/I417V	M V	Antimutator ^a , PAA-sensitive
T4 L412I	I	Weak antimutator ^a , sensitive to reduced dGTP ^a
T4 S411T/L412M	TM	Weak antimutator ^a , sensitive to reduced dGTP ^a
Sc DNA pol δ	608 D fn SLYPSIM	Wildtype
Sc L612M δ	M	Mutator ^b , PAA-sensitive ^b
Sc L612M/V758Mδ ^{e,f}	M	Wildtype fidelity, PAA-resistant
Sc DNA pol α	864 D fn SLYPSII	Wildtype
Sc L868M-α	M	Mutator ^c , PAA-sensitive ^b
Sc L868F-α	F	Mutator ^c , PAA-sensitive ^d
Sc L868W-α	W	Weak mutator ^c , PAA-resistant ^d

^aData from **Table 2** and Reha-Krantz and Nonay (1994).^bData from Li et al. (2005); mutator phenotype observed in the absence of mismatch repair.^cData from Niimi et al. (2004).^dUnpublished observations, LR-K.^eV758M is near Motif C.^fLi (2004).

1 mM PPi. Reactions with Mn²⁺ or Ca²⁺ ions also contained 15 mM Na citrate. The PPi concentration varied from 0.5 to 6 mM and Mn²⁺ or Ca²⁺ ion concentrations varied from 0.25 to 10 mM. Reactions were stopped by the addition of 2 μl 0.2 M EDTA. The product of pyrophosphorolysis, [α -³²P]dCTP, was separated by thin layer chromatography on polyethyleneimine impregnated cellulose (PEI) plates. Samples (2 μl) of the reactions were applied 1 cm from the bottom of the plate. The plates were developed in 50% ethanol up to 5 cm, dried and developed in 0.15 M KH₂PO₄ – 15% ethanol solvent until the solvent reached the top of the plate. Plates were dried and either the UV-absorbing spot was cut out and radioactivity was determined using a scintillation counter or the dried TLC plate was exposed to a PhosphorImager screen (Molecular Dynamics). These conditions were also studied under DNA sequencing conditions (see below).

Pyrophosphorolysis was also determined in assays using the fluorescence of the base analog, 2-aminopurine (2AP). The use of 2AP to study DNA polymerase function is described with detailed instruction by Reha-Krantz (2009). Briefly, 2AP fluorescence in DNA is quenched by base-stacking interactions, but the 2AP nucleotide is highly fluorescent. Thus, for primer-templates labeled at the 3'-end of a primer with 2AP, pyrophosphorolysis is detected as an increase in fluorescence intensity due to production of free 2AP deoxynucleoside triphosphate (d2APTP). Because 3'-exonuclease activity will also release the terminal 2AP nucleotide, but as the deoxynucleoside monophosphate (d2APMP), DNA polymerases were engineered to be exonuclease deficient with the D112A/E114A amino acid substitutions

which prevent Mg²⁺ binding in the exonuclease active site (Reha-Krantz and Nonay, 1993; Elisseeva et al., 1999). Stopped-flow experiments were performed with the Applied Photophysics SX.18 MV spectrofluorometer. Samples were excited at 310 nm; a 335 nm cut-off filter was used. Temperature in the sample-handling unit was maintained at 20.0 ± 0.5°C. Reactions were initiated in the stopped flow by mixing equal volumes of a solution of 1400 nM exonuclease-deficient T4 DNA polymerase, 400 nM 2AP-DNA substrate, 5 mM PPi, 25 mM HEPES (pH 7.5), 2 mM DTT, 50 mM NaCl and 1 mM EDTA with a second solution of 16 mM MgCl₂, 25 mM HEPES (pH 7.5), and 50 mM NaCl. After mixing, the final concentrations of reaction components were 700 nM exonuclease-deficient T4 DNA polymerase, 200 nM 2AP-DNA, 2.5 mM PPi, 8 mM MgCl₂, 1 mM DTT, 25 mM HEPES (pH 7.5), 50 mM NaCl, and 0.5 mM EDTA. Between five and six determinations were performed for each reaction and mean values were calculated. The experimental traces were fit to either single or double exponential equations. The agreement of the curve fits was judged by analysis of the randomness of the distribution of residuals for curves generated by single or double exponential equations. Two DNA substrates were used that differed in the A+T- or G+C-richness in the primer-terminal region. A+T DNA: primer, 5'GCACGTCATCGGTAA**T**P; template, 3'CGTCAGTAGCCATTATGGATCGATGGTTT. G+C DNA: primer, 5'GCACGTCATTAA**C**GGTP; template, 3'CGTGCAGTAATTGCCATGGATCGATGGTTT. The A+T and G+C primer terminal DNA sequences are indicated in bold font. P indicates 2AP.

Table 2 | DNA replication fidelity by phage T4 wild type and mutant polymerases with amino acid substitutions in motif A.

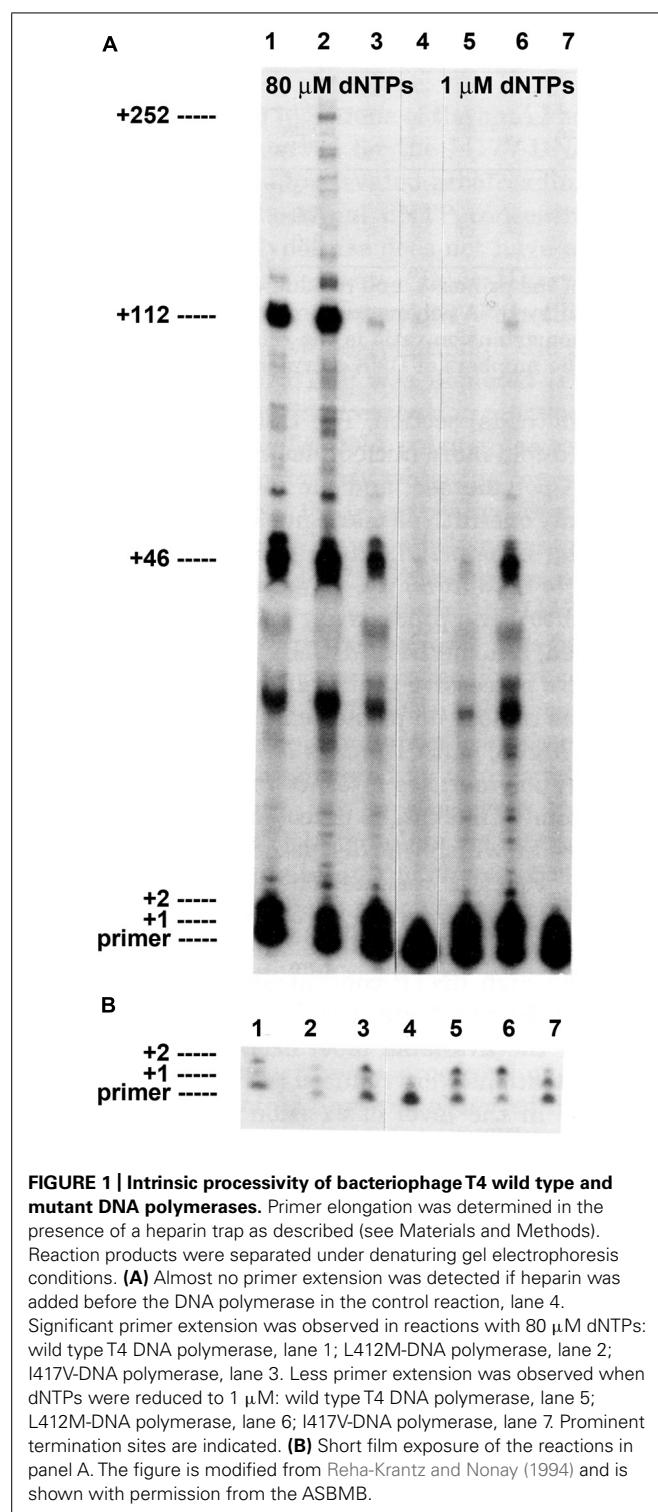
DNA polymerases	Forward mutation frequency ^a	Types of ac mutations (percentage)	Base substitution reversion frequency ^b	Frameshift reversion frequency ^c	
Wild type ^d	1/10 ⁵	GC→AT AT→GC GC→TA GC→CG AT→TA AT→CG Total base substitution Total frameshift Total other ^e	(18%) (12%) (1%) (3%) (5%) (1%) 40% 28% 32%	0.7/10 ⁶	1/10 ⁶
L412M ^d	35/10 ⁵	GC→AT AT→GC GC→TA GC→CG AT→TA AT→CG Total base substitution Total frameshift Total other	(37%) (2%) (13%) (2%) (8%) (9%) 90% 3% 7%	9/10 ⁶	3/10 ⁶
I417V	nd ^f	nd	0.005/10 ^{6,g}	nd	
L412M/I417V	nd	nd	0.03/10 ^{6,g}	nd	
L412I	nd	nd	0.1/10 ^{6,g}	nd	
S411T/L412M	nd	nd	0.07/10 ^{6,g}	nd	

^a Mutations that confer acriflavin resistance in the ac gene (Wang and Ripley, 1998).^b Reversion of rll199oc to rll; mostly AT→GC mutations (Reha-Krantz, 1995).^c Reversion of rll131 by +1 fs.^d Data from several students, especially Alia Daoud and Fatima Kamal.^e Total "other" are mostly deletions between direct repeats.^f nd; not determined.^g Data from Reha-Krantz and Nonay (1994).

DNA SEQUENCING WITH THE T4 EXONUCLEASE-DEFICIENT, L412M-DNA POLYMERASE HOLOENZYME COMPLEX: FIDELITY™

FIDELITY™ (ONCOR, Gaithersburg, MD, USA) was a manual DNA sequencing kit that was marketed for general sequencing purposes but also for use to sequence difficult DNA templates. There are four steps. In the first step, primer (1 pmol), labeled at the 5'-end in reactions using T4 polynucleotide kinase and [γ^{32} P]ATP, was annealed in a 10 μ l reaction to 1 μ g ssDNA M13 DNA in annealing buffer [25 mM Tris-HCl (pH 8.5), 20 mM MgCl₂, and 50 mM NaCl]. Reactions were held at 65°C for 2 min and then cooled slowly over 30 min to 35°C. The reactions were then pulse-centrifuged and chilled briefly on ice. If the primer was not labeled, a second step was performed to label the primer DNA internally using [α^{35} S] or [α^{33} P]dATP at 1500 Ci/mmol, 10 mCi/ml. Internal labeling reactions contained the 10 μ l annealed DNA produced in the first step; 3 μ l T4 reaction buffer [0.4 M Tris-HCl (pH 8.5), 40 mM MgCl₂, 40 mM DTT,

0.4 mg/ml acetylated BSA; 4 mM ATP; 1.4 μ M each dTTP, dGTP, dCTP]; exonuclease-deficient L412M-DNA polymerase (10 nM); and water to produce a final total volume of 18 μ l. The reaction was incubated at 40 to 42°C for 15 min and then placed on ice. Primer elongation with chain terminators was carried out in the third step. A mixture (6 μ l) of T4 DNA polymerase accessory proteins (gp32, 2.7 mg/ml; gp44/62, 0.2 mg/ml; gp45, 0.7 mg/ml) in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, and 62% glycerol was added to the 18 μ l of labeled DNA and mixed. 5.5 μ l of the elongation mix was added to 2 μ l of the four separate termination reaction tubes. All termination mixes contained 300 μ M dATP, dTTP, dGTP, and dCTP that were supplemented with 40 μ M 3'NH₂ddATP (A mix), or 80 μ M 3'NH₂ddTTP (T mix), or 40 μ M 3'NH₂ddGTP (G mix), or 80 μ M 3'NH₂ddCTP (C mix). The reactions were incubated at 40 to 42°C for 5 min. In the last (4th step), reactions were stopped and proteins degraded by addition of 5 μ l of a solution containing



3 parts Stop Solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol blue) and 1 part Proteinase K Solution (100 μ g/ml proteinase K in 50 mM Tris-HCl, pH 8.0). The reactions were mixed and incubated at 40°C for 15 min. Samples were heated at 80°C immediately prior to loading samples for gel electrophoresis.

Note, for reactions with inorganic pyrophosphatase (iPPase), the T4 DNA polymerase solution also contained 8 U/ml yeast iPPase. Alternatively, pyrophosphorolysis was inhibited in reactions with 16.7 mM Mg²⁺, 15 mM Na citrate and either 2.5 mM Mn²⁺ or 10 mM Ca²⁺.

RESULTS

BIOCHEMICAL CHARACTERIZATION OF THE BACTERIOPHAGE T4 L412M-DNA POLYMERASE

Increased intrinsic processivity

DNA polymerase intrinsic processivity refers to the number of nucleotides incorporated during one enzyme-DNA association replication cycle, but intrinsic processivity is partially masked *in vivo* by the association of processivity factors. For many DNA polymerases that function at replication forks, a clamp protein (gp45 in phage T4, PCNA in eukaryotes and the beta sliding clamp in *Escherichia coli*) tethers the DNA polymerase to the DNA template. The gp45 and PCNA clamps are composed of three identical subunits that form a doughnut structure around duplex DNA. The clamp is proposed to function as a tool belt to tether one or more DNA polymerases or other proteins found at replication forks (Pagès and Fuchs, 2002; Indiani et al., 2005; Moldovan et al., 2007). Although researchers have known for years that tethering of the T4 DNA polymerase in holoenzyme complexes increases both polymerase and exonuclease activities, Yang et al. (2004) observed that the primer-template appears to transfer between two tethered polymerases during proofreading. Thus, apparent continuous replication by the tethered T4 DNA polymerase may involve coordinated replication by two DNA polymerases tethered to the same clamp (Joyce, 2004). Tethering, however, does not fully compensate for differences in intrinsic processivity as explained below.

Without tethering, the wild type T4 DNA polymerase has little ability to initiate or extend a primer (Figure 1; Reha-Krantz and Nonay, 1993, 1994; Spacciapoli and Nossal, 1994). In reactions with ³²P-labeled primer, single-stranded circular DNA and heparin to trap dissociated DNA polymerase, the vast majority of primers are not extended by the wild type T4 DNA polymerase and the few extensions detected are limited to incorporation of one or two nucleotides (Figure 1B, lane 1). Longer extension products are visible with prolonged film exposure (Figure 1A, lane 1), which means that most non-tethered T4 DNA polymerase molecules cannot extend the primer, the few complexes that are competent for extension incorporate just one or two nucleotides before dissociation, and only a few of these complexes escape the initiation phase and commence primer elongation. However, for primer extension complexes that escape early dissociation, there are preferential termination sites as indicated by discrete bands after incorporation of 46, 112, and 252 nucleotides (Figure 1A). Thus, the intrinsic processivity of the wild type T4 DNA changes from an initial very low or essentially non-processive state to a state with higher processivity during the elongation phase, but the elongating DNA polymerase remains sensitive to difficult DNA sequences as demonstrated by preferential termination sites. The role of the clamp then is in initiation and in assisting replication through difficult DNA sequences.

Amino acid substitutions in Motif A of the T4 DNA polymerase (**Table 1**) dramatically alter intrinsic processivity. Longer extension products are visible for the L412M-DNA polymerase (**Figure 1A**, lane 2), especially when concentrations of deoxyribonucleoside triphosphates (dNTPs) are reduced to 1 μ M (**Figure 1**, lane 6); thus, more L412M-DNA polymerase complexes are able to escape the initiation phase and to replicate past difficult sequences. Kinetic studies are consistent with these observations. In standing start primer extension assays, the K_d for dAMP incorporation is lower for the L412M-DNA polymerase compared to wild type, 11 versus 16 μ M (Hariharan et al., 2006), and the k_{off} rate is about twofold slower (Fidalgo da Silva et al., 2002). While the L412M substitution is the most conservative substitution, dramatic changes in the opposite direction are observed for the slightly less conservative L412I substitution. The L412I-DNA polymerase cannot replicate DNA when the dGTP pool is reduced and an antimutator phenotype is observed instead of the mutator phenotype observed for the L412M-DNA polymerase (**Table 1**). These features resemble the consequences of another conservative substitution in Motif A, I417V. The I417V-DNA polymerase is much less processive than the wild type T4 DNA polymerase (**Figure 1**, lane 3), especially when dNTP pools are reduced (**Figure 1**, lane 7). Like the L412I-DNA polymerase, the I417V-DNA polymerase cannot replicate DNA when the dGTP pool is reduced (Reha-Krantz and Nonay, 1994; Stocki et al., 1995) and proofreading is increased as indicated by the antimutator phenotype (**Table 2**). The low processivity of the I417V-DNA polymerase means that replication is inhibited by subtle DNA damage, for example a phosphotriester, which does not impede replication by other DNA polymerases (Tsujiwaka et al., 2003). Because the differences in replication fidelity and sensitivity to dNTP pools are observed *in vivo* for the mutant DNA polymerases in the presence of the gp45 clamp, intrinsic processivity prevails in the presence of tethering and determines the stability of DNA polymerase complexes when dNTP pools are low, when DNA templates are damaged or difficult sequences are present, and if proofreading will be initiated.

Sequences at preferential termination sites are informative because they identify DNA sequences that are difficult for the T4 DNA polymerase to replicate. Many termination sites occur at simple repeats: the +46 product terminates within the template ATAT sequence and the +112 product terminates at the beginning of the template GCGC sequence (**Figure 1**; Reha-Krantz and Nonay, 1993). Termination with another DNA template was observed at the beginning of a template CACACA sequence (Spacciapoli and Nossal, 1994). These dinucleotide repeat termination sites are distinct from another type of difficult sequence, namely hairpin DNA structures that form between inverted repeats in ssDNA ahead of an advancing DNA polymerase. The T4 DNA polymerase has little ability to bypass hairpin structures in the absence of T4 ssDNA binding protein (gp32) and the clamp (gp45) and clamp loading proteins (Huang et al., 1981; Roth et al., 1982; Bedinger et al., 1989; Jarvis et al., 1991; Hacker and Alberts, 1994).

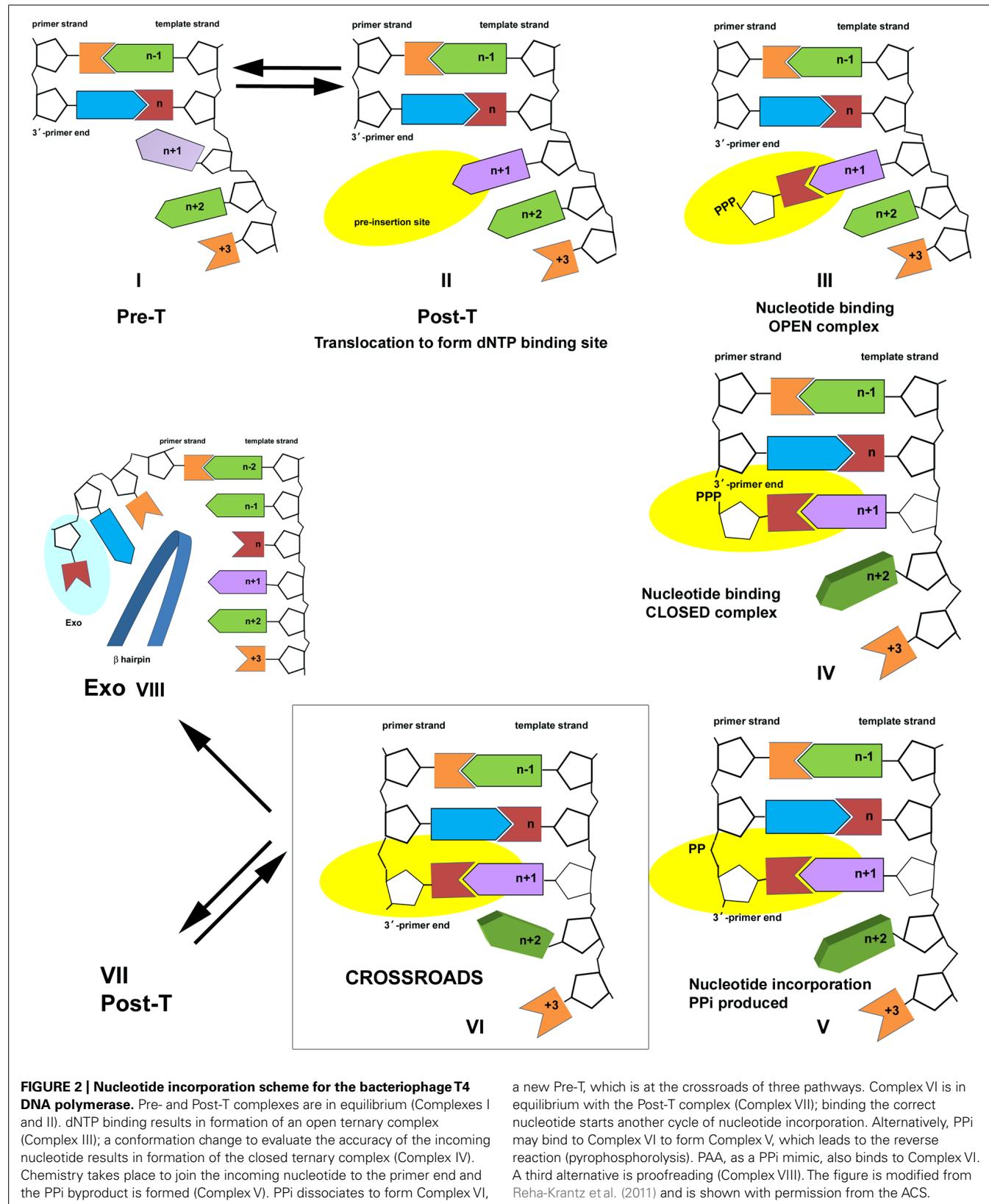
The simple repeat sequences at the termination sites in **Figure 1** likely cause pausing, which causes enzyme dissociation that is

associated with initiation of the proofreading pathway (see review by Reha-Krantz, 2010). While mismatches at the primer-end trigger proofreading as an error avoidance mechanism (Brutlag and Kornberg, 1972; Muzyczka et al., 1972), gratuitous proofreading of correctly paired primer ends can be caused by anything that hinders continued primer elongation. In contrast to dissociation of the T4 DNA polymerase when proofreading is initiated, return of the trimmed primer-end to the polymerase active site is rapid and processive, even in the absence of tethering (Reddy et al., 1992; Fidalgo da Silva and Reha-Krantz, 2007). With respect to *in vitro* applications, less gratuitous proofreading and subsequent dissociation is observed for the T4 L412M-DNA polymerase than for other DNA polymerases; hence, difficult template sequences, low dNTP concentrations, and other factors that impede nucleotide incorporation will be less problematic for the more processive DNA polymerase.

Biochemical mechanism for increased intrinsic processivity by the T4 L412M-DNA polymerase; insights provided by heightened sensitivity to PAA

The L412M substitution converts the T4 DNA polymerase from resistance to sensitivity to the herpes virus antiviral drug, phosphonoacetic acid (PAA). PAA-sensitivity is observed for the T4 L412M-DNA polymerase *in vivo* (**Table 1**; Reha-Krantz et al., 1993; Li et al., 2010) and *in vitro* (Reha-Krantz and Nonay, 1994). PAA resembles pyrophosphate (PPi), a byproduct of nucleotide incorporation; hence, PAA-sensitivity indicates increased PPi binding. PAA-sensitivity can be understood in terms of the nucleotide incorporation pathway described in **Figure 2** (Reha-Krantz et al., 2011). Complex I is a pre-translocation (Pre-T) complex. Complex II is formed by translocation forward by one template position, which removes the PPi binding site present in Complex I and creates a new nucleotide binding pocket (pre-insertion site). Complexes I and II are in rapid equilibrium (Fidalgo da Silva et al., 2002; Hariharan et al., 2006). Nucleotide binding produces the open ternary complex, Complex III, and a conformational change produces the closed ternary complex, Complex IV. Nucleotide incorporation takes place in Complex V; then PPi is released to form Complex VI, a Pre-T complex like Complex I except that the primer-end has been extended by one nucleotide.

Complex VI is at the crossroads of three competing pathways. The DNA polymerase may continue nucleotide incorporation by translocating to form Complex VII, a new post-T complex, or the reverse reaction can be initiated by re-binding PPi to form Complex V and then catalyzing pyrophosphorolysis, which produces dNTP and shortens the primer strand by one nucleotide (Complex IV). This is the pathway that is sensitive to PPi-like antiviral drugs. A third possibility is proofreading, which is initiated if the primer-end is mismatched, but gratuitous proofreading is also possible if polymerization conditions are not optimal, for example low dNTP pools. For proofreading, the end of the primer strand is separated from the template and transferred to the exonuclease site. A beta hairpin structure located in the exonuclease domain of the T4 DNA polymerase acts as a wedge to stabilize the exonuclease complex, Complex VIII (Stocki et al., 1995; Marquez and Reha-Krantz, 1996; Reha-Krantz, 1998, 2010; Reha-Krantz et al., 1998; Hogg



et al., 2004, 2007; Subuddhi et al., 2008). All three pathways are in competition, but differences in DNA polymerase interactions with the primer-template affect which pathway is chosen. PAA-sensitivity indicates that the T4 L412M-DNA polymerase favors forming the Pre-T complex, which has a PPi-binding site, over the Post-T complex, which has a dNTP binding site. Although PPi-like antiviral drugs could conceivably inhibit replication by interfering with dNTP binding in Post-T complexes, structural studies show that PPi-like phosphonoformic acid (PFA) traps a mutant RB69 DNA polymerase in a complex that resembles Complex V (Zahn et al., 2011) as expected if PFA binds to the Pre-T complex.

Formation of Pre- and Post-T complexes can also be observed in assays that use the fluorescence of the base analog 2AP. Pre-T complexes have higher fluorescence intensity than Post-T complexes for DNA substrates with 2AP positioned in the +1 position in the template strand, adjacent to the terminal base pair (Mandal et al., 2002; Hariharan and Reha-Krantz, 2005; Reha-Krantz et al., 2011). The higher fluorescence intensity of polymerase complexes formed with the L412M-DNA polymerase indicates that the L412M-DNA polymerase favors formation of the highly fluorescent Pre-T complexes over Post-T complexes in the absence of PPi and dNTPs (Hariharan and Reha-Krantz, 2005). The role of the Pre-T complex in drug sensitivity is further substantiated by studies that show that PFA inhibits the HIV-1 reverse transcriptase by trapping a Pre-T complex (Marchand et al., 2007). We extend these findings by linking PAA-sensitivity of the L412M-DNA polymerase to increased intrinsic processivity, which suggests that Pre-T complexes are more stable, and less subject to enzyme dissociation than Post-T complexes (see further discussion by Li et al., 2010).

Increased pyrophosphorolysis is observed for the T4 L412M-DNA polymerase

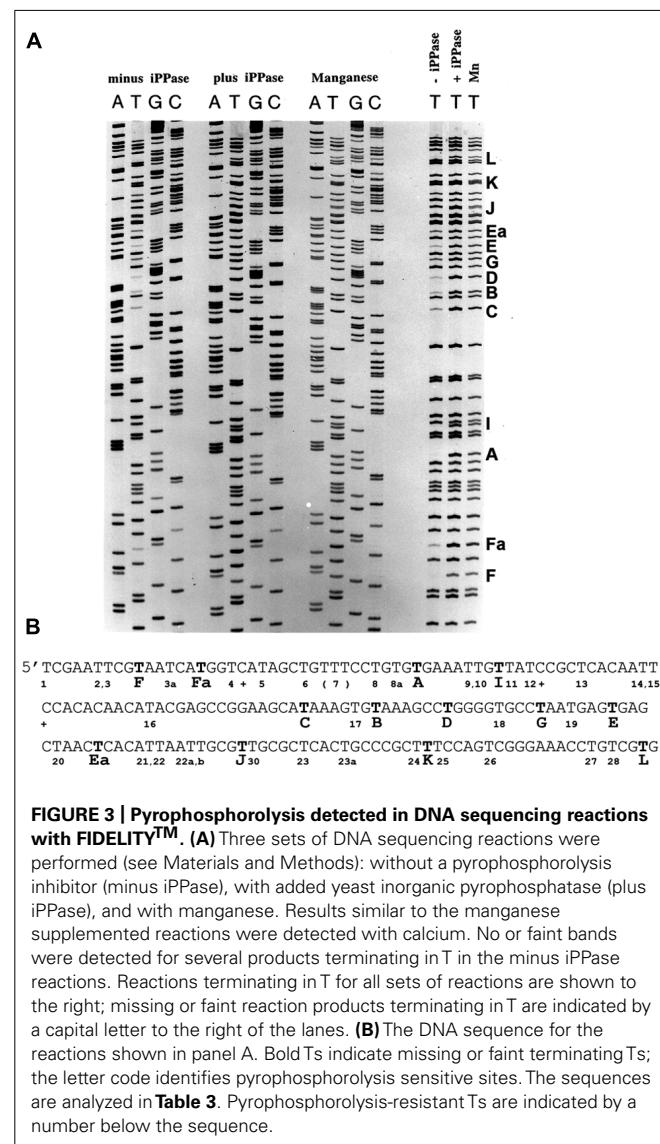
Increased pyrophosphorolysis is expected for PAA-sensitive DNA polymerases because the equilibrium between Pre- and Post-T complexes favors Pre-T complexes, which can bind PPi (Figure 2). The PAA-sensitive T4 L412M-DNA polymerase catalyzes a robust pyrophosphorolysis reaction. Section “Materials and Methods” for assay conditions. The pyrophosphorolysis rate in the presence of 2.5 mM PPi for the exonuclease-deficient (D112A/E114A) L412M-DNA polymerase is fivefold higher than the rate detected for the exonuclease-deficient T4 DNA polymerase with the wild type Motif A sequence and more than 80-fold higher than the less processive, exonuclease-deficient I417V-DNA polymerase. The K_m for PPi for the exonuclease-deficient L412M-DNA polymerase was ~ 0.5 mM compared to 2 mM for the exonuclease-deficient DNA polymerase with the wild type Motif A and was too high to be measured for the exonuclease-deficient I417V-DNA polymerase (Damaraju and Reha-Krantz, unpublished observations).

Mechanisms to curb pyrophosphorolysis catalyzed by the L412M-DNA polymerase; formation of divalent cation-PPi complexes

Pyrophosphorolysis is the reversal of the polymerization reaction, but pyrophosphorolysis is usually not a problem during DNA replication *in vivo* because iPPase degrades PPi to prevent build-up

of the high concentrations of PPi needed to drive pyrophosphorolysis. But high concentrations of PPi formed during DNA sequencing reactions support pyrophosphorolysis, which results in the degradation of chain-terminated sequencing products and ambiguous sequencing data (Figure 3). The addition of iPPase to sequencing reactions is often used to prevent pyrophosphorolysis (Tabor and Richardson, 1990), and this is successful in sequencing reactions catalyzed by the L412M-DNA polymerase, but divalent metal ions – Mn^{2+} and Ca^{2+} , are equally effective when combined with Mg^{2+} in the presence of 15 mM Na citrate (Figure 3; Damaraju and Reha-Krantz, unpublished observations).

The mechanism of inhibition of pyrophosphorolysis by divalent metal cations is complex because, as demonstrated for yeast iPPase (Ridlington and Butler, 1972), while some divalent metal ions such as Mg^{2+} are needed for catalysis (activators), other metal ions such as Mn^{2+} or Ca^{2+} may inhibit activity by binding to the enzyme or DNA or by forming complexes with PPi. We observed this complexity for the nucleotide incorporation and



pyrophosphorolysis reactions catalyzed by the T4 exonuclease-deficient L412M-DNA polymerase (Damaraju and Reha-Krantz, unpublished data). Either Mg^{2+} or Mn^{2+} can support nucleotide incorporation and pyrophosphorolysis reactions, but Mg^{2+} supports both reactions optimally at concentrations from 1 to >5 mM, while Mn^{2+} is only half as effective as Mg^{2+} and over a narrow concentration range that peaks at about 0.5 mM. In the presence of 15 mM Na citrate, however, Mn^{2+} supports nucleotide incorporation over a broad concentration range up to 5 mM, but the pyrophosphorolysis reaction is supported only over the narrow range from 0.5 to 1 mM. Thus, Na citrate protects the nucleotide incorporation reaction at high concentrations of Mn^{2+} , but not the pyrophosphorolysis reaction. This means that pyrophosphorolysis catalyzed DNA degradation can be avoided in DNA sequencing reactions by the addition of 15 mM Na citrate and 3 to 5 mM Mn^{2+} , but note further improvements discussed below.

Na citrate protects the nucleotide incorporation reaction by functioning as a weak chelator, which reversibly binds divalent metal cations; however, PPi also forms complexes with metal cations. This point is demonstrated by the observation that higher amounts of Mn^{2+} ions are needed to inhibit pyrophosphorolysis by the T4 L412M-DNA polymerase as the concentration of PPi is increased (Damaraju and Reha-Krantz). For example, in the presence of 1 mM PPi, ~ 1 mM Mn^{2+} produces 50% inhibition of pyrophosphorolysis; for 2 mM PPi, ~ 2 mM Mn^{2+} is needed for 50% inhibition; and for 4 mM PPi, >3 mM Mn^{2+} is required to achieve the same level of inhibition (Figure 4). The stoichiometry of the reaction is consistent with formation of Mn-PPi chelate complexes or in the case of Ca^{2+} , Ca-PPi chelate complexes. Thus, there is a complex equilibria involving one or more metal ions (Me): DNA pol-Me, citrate-Me, PPi-Me, free Me $^{2+}$, free DNA pol, free PPi, free Na citrate, and active and inactive DNA pol-Me-PPi complexes. While Mn-PPi binding may inactivate pyrophosphorolysis directly, this is not the most likely explanation for inhibition in reactions with Na citrate and Mn^{2+} above 3 mM because nucleotide incorporation is not inhibited; Mn-PPi binding would be expected to inhibit both reactions. Another explanation is that Mn^{2+} sequesters PPi and reduces the high concentration of PPi needed for the pyrophosphorolysis reaction, but there is sufficient free Mn^{2+} (~ 0.5 mM) to support, but not to inhibit, the nucleotide incorporation reaction. The ability of Na citrate to maintain low concentrations of Mn^{2+} has been proposed by others (Beckman et al., 1985; Tabor and Richardson, 1989, 1990), but in this case Mn^{2+} was thought to improve the evenness of band intensities (DNA sequencing products) by reducing discrimination against chain-terminating nucleotides. Ca^{2+} can also chelate PPi and effectively prevent pyrophosphorolysis (results not shown).

Although conditions were optimized for Mn^{2+} -dependent nucleotide incorporation and minimal pyrophosphorolysis, nucleotide incorporation was only 60% of what is observed in Mg^{2+} -dependent reactions and Ca^{2+} does not support the nucleotide incorporation reaction catalyzed by T4 DNA polymerase. Optimal nucleotide incorporation and minimal pyrophosphorolysis activity were obtained in mixed, divalent

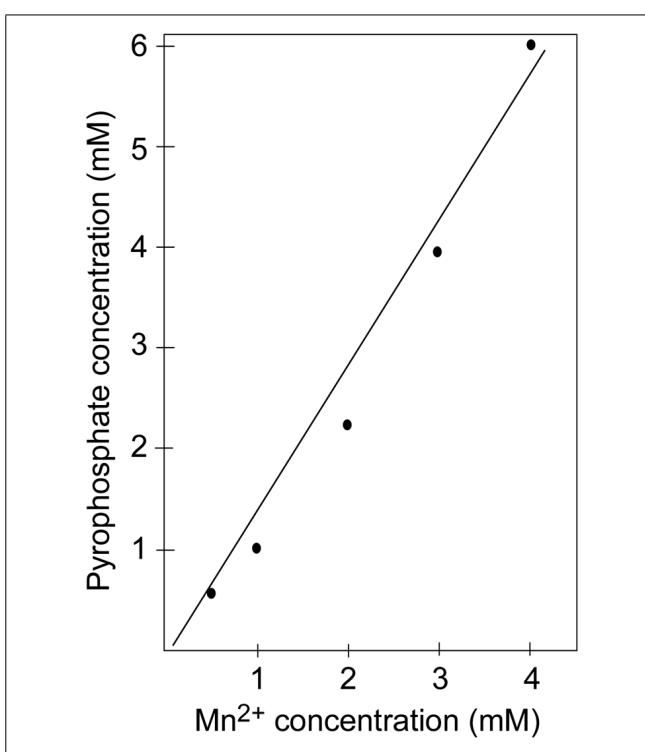


FIGURE 4 | Inhibiting pyrophosphorolysis with manganese. Increased manganese is needed to inhibit the pyrophosphorolysis reaction catalyzed by the exonuclease-deficient L412M-DNA polymerase when the concentration of PPi is increased.

metal reactions with Mg^{2+} and either Mn^{2+} or Ca^{2+} in the presence of Na citrate. For DNA sequencing using FIDELITY™ (see Materials and Methods), pyrophosphorolysis was essentially eliminated in reactions with 16.7 mM Mg^{2+} , 15 mM Na citrate and either 2.5 mM Mn^{2+} , or 10 mM Ca^{2+} without adverse effects on DNA sequencing efficiency (Figure 3).

DNA sequence context effects on the pyrophosphorolysis reaction catalyzed by the T4 L412M-DNA polymerase

DNA sequence effects have been reported for the pyrophosphorolysis reactions catalyzed by the Klenow fragment of *E. coli* DNA polymerase I (Mizrahi et al., 1986) and the phage T7 DNA polymerase (Tabor and Richardson, 1990), but systematic studies were not carried out to determine if specific DNA sequences are required for pyrophosphorolysis. We observed both general and specific DNA sequence contexts for the pyrophosphorolysis reactions catalyzed by the exonuclease-deficient T4 wild type and L412M-DNA polymerases.

Conditions that favor formation of Pre-T complexes (Complex VI, Figure 2) are predicted to increase pyrophosphorolysis because PPi can bind to this complex. In previous studies using the fluorescence of the base analog, 2AP, more Pre-T than Post-T complexes are formed with A+T-rich than G+C-rich DNA substrates (Hariharan and Reha-Krantz, 2005; Hariharan et al., 2006). Thus, pyrophosphorolysis is predicted to favor DNA substrates with A+T-rich primer-terminal regions. This was observed in reactions with A+T- and G+C-rich DNAs labeled

with 2AP at the 3'-primer end (see Materials and Methods). The rate for release of the terminal d2APTP by the exonuclease-deficient L412M-DNA polymerase in the presence of 2.5 mM PPi was sixfold faster for the A+T-rich DNA compared to the G+C-rich DNA, 14 s⁻¹ compared to 2.4 s⁻¹ (Damaraju and Reha-Krantz). The exonuclease-deficient T4 DNA polymerase with the wild type Motif A sequence also favored A+T-rich DNA, but with slower rates, 3 and 0.5 s⁻¹ for A+T- and G+C-rich DNAs, respectively. However, while pyrophosphorolysis was observed for DNA substrates with a terminal 2AP opposite template T, pyrophosphorolysis was not detected for 2AP opposite template C as expected if pyrophosphorolysis requires a matched primer-end. DNA substrates with A+T-rich primer-terminal regions also favor the exonucleolytic proofreading pathway (Bessman and Reha-Krantz, 1977; Bloom et al., 1994; Reha-Krantz, 2010). Thus, increased formation of Pre-T complexes, which likely reflects decreased ability to form Post-T complexes, increases the likelihood of proofreading or pyrophosphorolysis. Proofreading is favored if the primer-end is mismatched and pyrophosphorolysis is observed if sufficient PPi is present.

Specific DNA sequences were observed to promote pyrophosphorolysis under DNA sequencing conditions using the FIDELITY™ manual DNA sequencing kit developed by ONCOR, Gaithersburg, MD, USA. Reactions contained the exonuclease deficient L412M-DNA polymerase, the gp45 processivity factor, the gp45 loading proteins as well as the T4 ssDNA binding protein, gp32, to prevent formation of DNA hairpin structures in ssDNA. Reaction products were internally labeled with α^{33} P-dATP; the template was single-stranded circular DNA. Instead of 2',3'-dideoxy nucleoside triphosphate terminators (ddNTPs), 3'-amino-2',3'-dideoxy nucleoside triphosphates (3'NH₂ddNTPs) were used because the T4 DNA polymerase showed less discrimination. 3'NH₂ddNTPs can be purchased from TriLink BioTechnologies, San Diego, CA, USA. All reaction components and procedures are described in Section "Materials and Methods." Four separate reactions were run, each with a different chain-terminator. The reaction products were separated by polyacrylamide gel electrophoresis under denaturing conditions (Figure 3A). In the absence of agents that reduce pyrophosphorolysis—iPPase or manganese (Mn^{2+}), there are missing bands due to pyrophosphorolysis as well as many faint bands (Figures 3A,B).

It is readily apparent that not all chain-terminated primer termini are equally sensitive to pyrophosphorolysis; only primer ends ending in "T" were subject to pyrophosphorolysis, but not all terminal Ts (see right side of Figure 3A). Examination of 200 nucleotides of DNA sequence (Figure 3B) with 54 sequencing products that terminated in T revealed that only 12 of the terminal Ts were subject to pyrophosphorolysis. What is different about the DNA sequences for pyrophosphorolysis-sensitive and insensitive sites? There are three related sequences in the primer-template region that promote pyrophosphorolysis (Table 3); the sequences differ only in the base pairs at the -1 and -2 positions. Sequence 1 has a G_{template}C_{primer} base pair in the -2 position; sequence 2 has a C_{template}G_{primer} base pair in the -1 position, but not in the -2 position; and sequence 1/2 has a G_{template}C_{primer} base pair in the -2 and

a C_{template}G_{primer} base pair in the -1 position. A template "G" was never observed in the +1 position in the template strand, even though statistically 2 or 3 were expected in the 11 sequences examined. Note that a "G" in the +1 position for a terminal T at position 26 in the DNA sequence (Figure 3B) stabilized the T to pyrophosphorolysis even though the remainder of the consensus sequence is present. Similarly, a template "A" was not observed in the +4 position, but an A stabilized the Ts at DNA sequence positions 5 and 8. The consensus sequences 1, 2 and 1/2 (Table 3) are required to sensitize terminal Ts to pyrophosphorolysis, but one exception was observed, see Ea in Figure 3. Although the consensus sequence is not present, the terminal T is just before a dinucleotide GT repeat, which may stall formation of Post T complexes and, thus, promote pyrophosphorolysis. Thus, the pyrophosphorolysis sensitive sites observed in Figure 3 may be an underestimation of the number of sensitive sites. Nevertheless, the strong bias for pyrophosphorolysis for only primers ending in "T" and only for a subset of terminal "Ts" indicates sequence specificity for the pyrophosphorolysis reaction catalyzed by the T4 L412M-DNA polymerase.

Increased incorporation of fluorophore-labeled nucleotides

The L412M-DNA polymerase has increased ability to incorporate modified nucleotides compared to the wild type T4 DNA polymerase, exonuclease-deficient T4 DNA polymerases and other mutant T4 DNA polymerases and DNA polymerases from other organisms (Goodman and Reha-Krantz, 1999). Reactions showing incorporation and extension of a variety of modified nucleotides with the L412M- and exonuclease-deficient L412M-DNA polymerase are shown in US patent 5945312. The modified nucleotides tested include, but are not limited to rhodamine-dUTP, fluorescein-dUTP, rhodamine-dCTP, biotin-dCTP and DIG-dCTP. For all modified nucleotides, the L412M-DNA polymerase performed significantly better than the wild type T4 DNA polymerase or DNA polymerases from other organisms. The T4 L412M-DNA polymerase was used to prepare fluorescently labeled DNA with one or two fluorophore-labeled bases for proof-in-principle testing of DNA sequencing by exonuclease digestion proposed by Keller and colleagues (Goodwin et al., 1995; Werner et al., 2003).

THE PHAGE T4 EXONUCLEASE-DEFICIENT L412M-DNA POLYMERASE AS A DNA SEQUENCING ENZYME: FIDELITY™

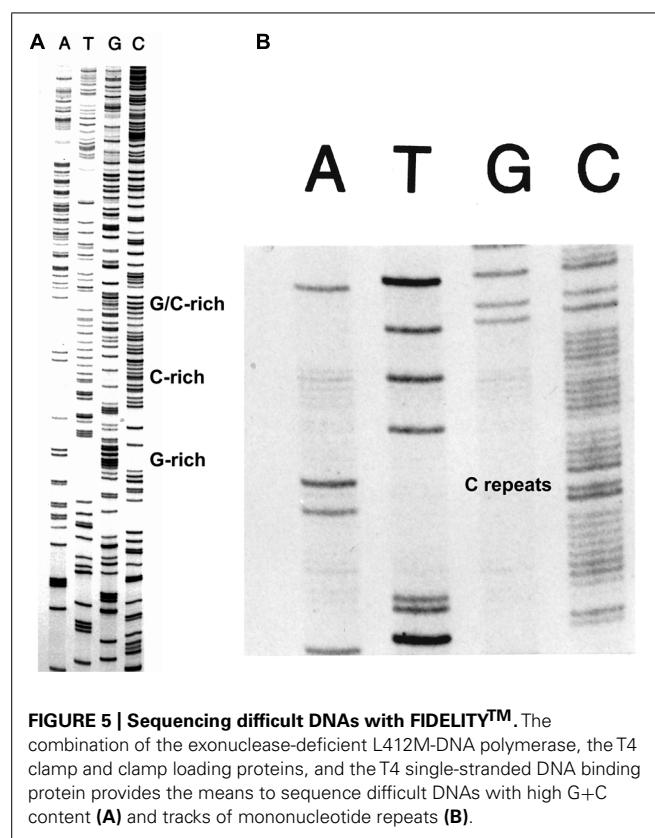
The FIDELITY™ DNA sequencing kit (ONCOR, Gaithersburg, MD, USA) was developed for routine DNA sequencing, but especially for sequencing difficult DNA sequences. The use of the exonuclease-deficient L412M-DNA polymerase with increased intrinsic processivity plus the addition of T4 processivity factors and ssDNA binding protein provided the means to sequence all types of difficult DNA sequences with greater success than achieved by other methods as judged by testimonials and applications, for example, see Devireddy and Jones (1998) and Tullius et al. (2001). An example of routine DNA sequencing is shown in Figure 3 and examples of sequencing difficult DNAs are shown in Figure 5. Note the clean sequence for the G+C-rich template

Table 3 | DNA sequences that promote pyrophosphorolysis by the phage T4 L412M-DNA polymerase.

Sequence	Primer-template									
	-4	-3	-2	-1	n ^a	+1	+2	+3	+4	
1	Template	X ^b	A/C	G	X	A	Not G	X	X	Not A
	Primer	X	T/G	C	X	T ^c				
	Fa^d	T A	A T	G C	T A	A T	C	C	A	G
	C	T A	C G	G C	T A	A T	T	T	T	C
	D	A T	C G	G C	G C	A T	C	C	C	C
	G	A T	C G	G C	G C	A T	T	T	A	C
	K	G C	C G	G C	A T	A T	A	G	G	T
2	Template	X	A/C	A/T	C	A	Not G	X	X	Not A
	Primer	X	T/G	T/A	G	T				
	A	A T	C G	A T	C G	A T	C	T	T	T
	I	T A	A T	A T	C G	A T	A	T	A	G
	B	T A	C G	A T	C G	A T	T	T	T	C
	E	A T	C G	T A	C G	A T	C	T	C	G
1/2	Template	X	A/C	G	C	A	Not G	X	X	Not A
	Primer	X	T/G	C	G	T				
	F	A T	A T	G C	C G	A T	T	T	A	G
	J	A T	C G	G C	C G	A T	C	G	C	G
	L	C G	A T	G C	C G	A T	C	G	G	T
Exception	Ea	A T	T A	T A	G C	A T	G	T	G	T

^an, terminal base pair.^bX is any nucleotide; A, C, G, T.^cChain terminator.^dLetter code that corresponds to faint or missing Ts in the sequencing reactions shown in **Figure 3**.

Bold indicates highly conserved sequence.



(Figure 5A) and for a template with several tracks of C repeats (Figure 5B).

USING GENETIC APPROACHES TO IDENTIFY “BIOTECH” DNA POLYMERASES: *IN VIVO* AND *IN VITRO* SELECTION METHODS

Genetic selection methods are powerful tools to identify mutant DNA polymerases on the basis of phenotype without the need to have a complete understanding of structure or function. The advantage of genetic selection strategies is that if the stringency is sufficiently strong, then only mutants with the desired phenotype will survive. This expedites “finding a needle in a haystack.” The phage T4 L412M-DNA polymerase was the result of two consecutive genetic selection schemes.

First, a genetic selection was carried out for the identification of mutant T4 DNA polymerases with a strong mutator phenotype. Details for the selection of mutator T4 DNA polymerases (Reha-Krantz et al., 1986; Reha-Krantz, 1988) and a similar strategy for mutator yeast DNA polymerase δ mutants (Murphy et al., 2006) are described. Several mutator T4 DNA polymerases were identified, but not the L412M-DNA polymerase because this mutant replicates DNA with relatively high fidelity as discussed (Table 2). For many isolates there was more than one mutation in the DNA polymerase gene. Single mutant strains had to be constructed in order to determine if one or more mutations were required to confer the mutator phenotype and to rule out contributing mutations in other genes. For one isolate, called *mel5*, there were 11 mutations in the DNA polymerase gene,

but only the mutation encoding the D131G amino acid substitution was required for mutator activity (Reha-Krantz, 1988). The D131G-DNA polymerase has strongly reduced ability to form exonuclease complexes and, as a consequence, has severely reduced 3'—5' exonuclease activity (Baker and Reha-Krantz, 1998).

Why were there so many mutations in the *mel5* strain? One possibility is that error-prone replication by mutator DNA polymerases created many mutations within the DNA polymerase gene and in other genes during propagation, but a more likely explanation is that additional DNA polymerase mutations were selected to temper the mutational burden produced by the D131G amino acid substitution. This proposal is supported by the presence of a mutation in the *mel5* strain that encodes the I417V substitution in Motif A (Table 1). The I417V-DNA polymerase has an antimutator phenotype, increased proofreading, and low intrinsic processivity (Reha-Krantz and Nonay, 1994; Figure 1 and Table 2), activities which would help to maximize residual exonuclease activity in the presence of the D131G substitution and increase replication fidelity. These activities, however, would also help to improve maturation of Okazaki fragments. Okazaki fragment maturation requires the coordinated action of DNA polymerase 3'—5' exonuclease activity with a 5'—3' exonuclease, ExoI/RNaseH in T4 and Rad27 in yeast, to create a nick that can be sealed by DNA ligase (Jin et al., 2004; reviewed by Reha-Krantz, 2010). In the absence of DNA polymerase 3'—5' exonuclease activity, the DNA polymerase catalyzes strong displacement synthesis at junctions between Okazaki fragments that creates 5'—3' flaps that prevent ligation unless the flap is removed. Persisting unjoined Okazaki fragments are dangerous because strand discontinuities result in double strand breaks in the next replication cycle. Thus, there are at least two strong selective pressures exerted to decrease the negative effects of reduced DNA polymerase proofreading.

The second selection to identify the L412M-DNA polymerase was to select for suppressors of the low processivity and high proofreading activity of the I417V-DNA polymerase (Table 1 and Figure 1). One consequence of low processivity is the need for high dNTP pools to sustain replication. The I417V-DNA polymerase could not replicate DNA when the dGTP pool was reduced, which allowed for selection of second site mutations that allowed replication under low dGTP conditions (Reha-Krantz and Nonay, 1994; Stocki et al., 1995; Li et al., 2010). Mutations encoding the L412M substitution in Motif A and elsewhere in the DNA polymerase gene were identified (Stocki et al., 1995; Li et al., 2010). Thus, the mutation encoding the L412M substitution and functionally equivalent amino acid substitutions are expected to be identified as suppressors of mutant DNA polymerases with low processivity.

Selection schemes can also be done *in vitro* to identify mutant DNA polymerases with desired properties for biotech applications. Bourn et al. (2011) describe an *in vitro* strategy beginning with a chimeric DNA polymerase derived from *Pyrococcus kodakaraensis* and *Pyrococcus furiosus*, which was named Kofu DNA polymerase. Several mutant DNA polymerases with increased processivity

and other desirable features were identified, but many had multiple mutations as we observed for *in vivo* selections with the T4 DNA polymerase. There are no reports to link amino acid substitutions in the Kofu DNA polymerase with specific polymerase properties and the amino acid substitutions in the DNA polymerase marketed as KAPA HiFi DNA polymerase by KAPA Biosystems Ltd., South Africa, are proprietary. Bourn et al. (2011) suggest, however, that advantageous properties may require the combined action of several amino acid substitutions. None of the amino acids disclosed, however, are in Motif A, but there could be amino acid substitutions in other DNA polymerase regions that affect processivity, for example see Stocki et al. (1995), Reha-Krantz and Wong (1996), and Li et al. (2010). While our genetic selections were done *in vivo*, *in vitro* strategies similar to Bourn et al. (2011) can be envisioned that exploit sensitivities to dNTP pool concentrations, PAA or PPi can be envisioned.

We have applied information learned about T4 DNA polymerase function to other DNA polymerases, namely yeast DNA polymerases alpha and delta. Amino acid changes in the yeast DNA polymerases that are analogous to the L412M substitution in the T4 DNA polymerase produce PAA-sensitivity (Table 1; Li et al., 2005). Because PAA-sensitivity is correlated with increased intrinsic processivity, the mutant yeast DNA polymerases are also predicted to have increased processivity; however, the mutant yeast DNA polymerases are more error prone than the L412M-DNA polymerase (Table 2; Niimi et al., 2004; Li et al., 2005). The differences in replication fidelity for the T4 and yeast DNA polymerases suggest differences in the equilibria between Pre- and Post-T complexes and in forming exonuclease complexes. However, fine tuning this balance can be achieved by testing other amino acid substitutions in Motif A and by combining amino acid substitutions. For example, the phage T4 L412M/I417V-DNA polymerase is PAA-sensitive, although less than the L412M-DNA polymerase, but the double mutant has an antimutator phenotype and slightly increased proofreading compared to the wild type T4 DNA polymerase (Tables 1 and 2). Thus, single or multiple amino acid changes within the Motif A sequence of DNA polymerases in general may produce mutant DNA polymerases with the optimal balance of intrinsic processivity and proofreading that will allow replication of difficult DNA sequences and increased utilization of modified nucleotides while maintaining replication fidelity.

DISCUSSION

The phage T4 exonuclease-deficient L412M-DNA polymerase is proven to be an excellent DNA sequencing enzyme especially when combined with processivity factors (clamp and clamp loaders) and ssDNA binding protein; this is the basis of FIDELITYTM. All DNAs tested, including DNAs with long stretches of mono- and di-nucleotide repeats and high A+T- or G+C-content are sequenced cleanly (Figure 5; S. Woodgate personal observations). While the T4 L412M-DNA polymerase remains an attractive candidate as a DNA sequencing enzyme, other DNA polymerases may have additional desirable properties that would profit by modifications to the Motif A sequence. Indeed, in addition to our studies of Motif A in phage T4 and yeast DNA polymerases (Reha-Krantz and Nonay, 1994; Li et al., 2005; Table 1), Motif A in family B and A

DNA polymerases is surprisingly amenable to engineering despite being a conserved sequence (Table 1; Patel and Loeb, 2000; Niimi et al., 2004; Venkatesan et al., 2006, 2007). Our studies demonstrate that Motif A functions in determining intrinsic processivity for the phage T4 DNA polymerase (Figure 1) and, by extrapolation, to other DNA polymerases. The level of intrinsic processivity determines the ability of DNA polymerases to replicate difficult DNA sequences and to incorporate modified nucleotides; these properties are maintained even when dNTP pools are reduced for the L412M-DNA polymerase (Figure 1). Different amino acid substitutions and combinations of substitutions in Motif A allow fine-tuning to achieve the optimal balance of intrinsic processivity for the replication of difficult DNA templates while maintaining high replication fidelity; these are highly desirable properties for sequencing, amplification and labeling technologies. We focus here on Motif A because this motif is conserved in all DNA and RNA polymerases, which means that site-directed mutagenesis of Motif A can be used to identify mutant DNA polymerases with increased intrinsic processivity that still retain adequate proof-reading activity. However, other DNA polymerase regions also affect intrinsic processivity (Reha-Krantz et al., 1993; Stocki et al., 1995; Li et al., 2010) and, thus, are attractive secondary mutational targets.

Increased intrinsic processivity increases pyrophosphorylation, but methods were shown to curb pyrophosphorylation activity of the exonuclease-deficient L412M-DNA polymerase by the addition of iPPase or divalent metal ions, Mn²⁺ or Ca²⁺ (Figure 3). In the absence of these agents, pyrophosphorylation under DNA sequencing conditions displayed sequence specificity. Pyrophosphorylation was detected only for a subset of sequencing products terminated with T (Figure 3). Although pyrophosphorylation by the T4 wild type and L412M-DNA polymerases was higher for DNA substrates with A+T- compared to G+C-rich primer terminal regions, the sequence specificity observed under DNA sequencing conditions was specific for primer-ends with the T-terminating nucleotide plus either a C_{template}G_{primer} base pair in the -1 position or a G_{template}C_{primer} base pair in the -2 position or a C_{template}G_{primer} in the -1 and a G_{template}C_{primer} in the -2 position (Table 3). Dinucleotide repeats may also sensitize terminal Ts, see the pyrophosphorylation sensitive site Ea in Figure 3 and Table 3. The preference of pyrophosphorylation for A+T-rich primer-terminal regions is expected for the T4 DNA polymerase, because A+T-rich primer-termini favor formation of Pre-T complexes (Hariharan and Reha-Krantz, 2005), which can bind PPi and PPi-like antiviral drugs (Figure 2). The sequence-dependence for pyrophosphorylation under DNA sequencing conditions must also indicate sequences that favor Pre-T complexes, but the mechanism is unclear. Unfortunately, while there are numerous structural studies of the RB69 DNA polymerase (for example Franklin et al., 2001; Hogg et al., 2004; Wang et al., 2011), which is a close relative of the phage T4 DNA polymerase (Hogg et al., 2006), the primer-template DNAs used in crystallography did not have the pyrophosphorylation-sensitive sequences indicated in Table 3; thus, a direct test of DNA polymerase interactions with the pyrophosphorylation sequences has not been done. However, the DNA substrate used to capture a mutant RB69 DNA polymerase with the PPi analog, PFA (foscarnet), was A+T-rich except

for the terminal acyclo guanine (Zahn et al., 2011), which is consistent with our data that demonstrate increased pyrophosphorolysis with A+T-rich DNAs.

We propose that the sequence specificity for the pyrophosphorolysis reaction under DNA sequencing conditions is caused by subtle interactions of the DNA polymerase with a dynamic primer-terminal region that breathes more or less depending on the DNA sequence, but we do not rule out the possibility of base recognition, especially for bases in the +1 and +4 positions in the template strand (**Table 3**). Breathing in the terminal region is proposed to explain why A+T-richness favors proofreading by increasing the strand separation needed to form exonuclease complexes (for example, see Bloom et al., 1994 and Bessman and Reha-Krantz, 1977); we suggest here that breathing also affects the rapid equilibrium between Pre- and Post-T complexes and increased breathing favors Pre-T complexes. Breathing has been detected at the primer-terminal junction by fluorescence studies that demonstrate measurable unwinding ~ 2 base pairs into the duplex region (Jose et al., 2009), which may be exacerbated by DNA polymerase interactions. But how does the DNA polymerase distinguish between opting for pyrophosphorolysis compared to proofreading? Proofreading a chain-terminated primer-end may always be preferred, but the DNA polymerase in sequencing reactions is exonuclease-deficient which allows detection of pyrophosphorolysis, but pyrophosphorolysis is not simply a default pathway because sequence specificity is observed. The pyrophosphorolysis-sensitive DNA sequences (**Table 3**) must stabilize formation of stable Pre-T complexes (reduce dissociation) at the expense of Post-T complexes. In other words, sequences that promote pyrophosphorolysis must hinder translocation to form Post-T complexes.

A critical step in translocation may involve interactions of the DNA polymerase at the -1 position in the primer-terminal region, using the naming system indicated in **Table 1** with "n" designating the terminal base pair. Structural studies show that a conserved lysine residue intercalates into the primer-terminal region at the -1 position of ternary complexes formed with the RB69 DNA polymerase (Franklin et al., 2001), but in complexes trapped with the PPi mimic PFA (Zahn et al., 2011), translocation was blocked after nucleotide incorporation and lysine intercalation did not change, which meant intercalation was observed at the -2 position. Thus, lysine intercalation advances step wise after each nucleotide is incorporated with intercalation at the -2 base pair position in Pre-T complexes before translocation and advancing to the -1 base pair position to form Post-T complexes. The -1 and -2 positions are also critical for promoting pyrophosphorolysis; a G_{template}C_{primer} base pair is observed in the -2 position of sequences that promote pyrophosphorolysis (sequence 1) unless there is a C_{template}G_{primer} base pair in the -1 position (sequence 2) or two GC and CG base pairs are observed in the -2 and -1 positions, respectively (**Table 3**). If localized breathing in the primer-terminal region is the critical factor for forming stable Pre-T complexes, disfavoring exonuclease complexes, and hindering translocation to form Post-T complexes, then the requirement of GC and CG base pairs at the -2 and -1 positions suggests that GC base pairs impart stability that reduces breathing and strand separation needed to form exonuclease complexes. But what hinders

formation of Post-T complexes? One possibility may reflect the effect of duplex stability on the ability of the conserved lysine residue to intercalate into duplex DNA if the ease of intercalation is affected by A+T/G+C-richness. Another possibility may be that sequences which promote pyrophosphorolysis cause a slight distortion of the DNA helix. This possibility is suggested because not any GC base pair is sufficient, but only a G_{template}C_{primer} base pair in the -2 position or a C_{template}G_{primer} base pair in the -1 position, not the reverse Watson-Crick base pairs. There is also specificity for A_{template}T_{primer} or C_{template}G_{primer} base pairs in the -3 position. While base recognition has not been ruled out, subtle physical changes in the primer-terminal region appear to affect the equilibria between Pre- and Post-T complexes and exonuclease complexes. This suggestion is in keeping with trying to understand the underlying mechanism to explain how the conservative L412M substitution in the T4 DNA polymerase has such profound effects on processivity and sensitivity to the PPi-mimic PAA (**Figure 1**; Reha-Krantz and Nonay, 1994) while the L412I substitution also has profound effects, but in the opposite direction. Structural studies of the analogous RB69 L415M-DNA polymerase of Post-T ternary complexes with dNTP (Xie et al., 2013) or Pre-T complexes trapped with PFA (Zahn et al., 2011) do not shed light on mechanism, which then leaves open the possibility that rapidly changing conformations not yet captured in structural studies are behind the observed characteristics of the T4 L412M- and L412I-DNA polymerases.

Future studies are needed to determine if the pyrophosphorolysis-sensitive sequences described in **Table 3** promote pyrophosphorolysis in the absence of a chain-terminating T or if the chain-terminator contributes to the observed sequence specificity. DNA binding studies with the primer-templates described in **Table 3** are also needed. Studies are also needed to explore the possibility that bases in the +1 and +4 positions in the template are recognized by the T4 DNA polymerase. The ability of "A" in the +4 position in the template strand to prevent pyrophosphorolysis suggests that the DNA polymerase can detect sequence. Interestingly, the Pfu DNA polymerase detects deaminated bases in the +4 position as part of an apparent error avoidance mechanism (Connolly, 2009). The sequence specificity of the pyrophosphorolysis reaction suggests possible implications for antiviral drug therapy. For example, is the sensitivity of herpes viral DNA polymerases to PPi analogs, PAA/PFA/foscarnet, dependent on DNA sequence? See additional discussion by Li et al. (2010).

In conclusion, amino acid substitutions in the phage T4 DNA polymerase Motif A increase or decrease intrinsic processivity by altering the equilibria between Pre- and Post-T complexes and formation of exonuclease complexes. Amino acid changes in Motif A affect intrinsic processivity by increasing the stability of Pre-T complexes. Increased stability of Pre-T complexes means increased ability to replicate difficult DNA sequences and to incorporate modified nucleotides. Motif A is an attractive target for engineering "biotech" DNA polymerases because intrinsic processivity can be fine-tuned by using different amino acid substitutions and combinations of amino acid substitutions to optimize replication of difficult DNA sequences and to enhance the ability to use modified nucleotides while maintaining replication fidelity. Another important take-home message is to acknowledge that DNA and

DNA polymerase-DNA interactions are dynamic, especially in the primer-terminal region (Hariharan and Reha-Krantz, 2005; Hariharan et al., 2006; Jose et al., 2009). Lastly, DNA polymerases may have more ability to detect specific sequences than previously recognized.

AUTHOR CONTRIBUTIONS

Linda J. Reha-Krantz was responsible for organizing and writing the initial manuscript and for providing new data. Sandra Woodgate and Myron F. Goodman added intellectual content and critically evaluated the manuscript. Sandra Woodgate developed the FIDELITY™ DNA sequencing kit in collaboration with Myron F. Goodman and Linda J. Reha-Krantz and she provided data on the pyrophosphorolysis reaction under DNA sequencing conditions. All authors agree to be accountable for all aspects of the work presented.

ACKNOWLEDGMENTS

Linda J. Reha-Krantz acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) for operating and strategic grants, and from the Canadian National Cancer Institute. Linda J. Reha-Krantz also gratefully acknowledges the contributions of more than 40 skilled research personnel, students, and postdoctoral fellows from 1982 to the present. Myron F. Goodman has received support from NIHGM, NIHEHS and NIHCA grants. Sandra Woodgate acknowledges Jay George, CSO at ONCOR.

Selection of the mutant L412M-DNA polymerase and other mutant DNA T4 DNA polymerases, biochemical characterization of the mutant L412M-DNA polymerase and other mutant DNA T4 DNA polymerases, and determining the ability of the mutant L412M-DNA polymerase and other mutant T4 DNA polymerases to incorporate modified nucleotides are covered in the following U.S. patents authored by Myron F. Goodman and Linda J. Reha-Krantz: chain-terminating nucleotides for DNA sequencing, U.S. Patent Number – 5,547,859, Issued Aug. 20, 1996; Methods for identifying and isolating variant T4 DNA polymerases, U.S. Patent Number – 5,660,980, Issued Aug. 26, 1997; Variant DNA Polymerases, U.S. Patent Number – 5,928,919, Issued July 27, 1999; Synthesis of fluorophore-labeled DNA, U.S. Patent Number – 5,945,312, Issued Aug. 31, 1999. A licensing agreement was formed between ONCOR of Gaithersburg, Maryland and Myron F. Goodman and the University of Southern California and Linda J. Reha-Krantz and the University of Alberta. The FIDELITY™ kit was developed by Sandra Woodgate at ONCOR, Gaithersburg, Maryland in collaboration with Myron F. Goodman, Linda J. Reha-Krantz, and other researchers at ONCOR. The kit is no longer manufactured.

REFERENCES

- Arezi, B., McKinney, N., Hansen, C., Cayouette, M., Fox, J., Chen, K., et al. (2014). Compartmentalized self-replication under fast PCR cycling conditions yields Taq DNA polymerase mutants with increased DNA-binding affinity and blood resistance. *Front. Microbiol.* 5:408. doi: 10.3389/fmicb.2014.00408
- Baker, R., and Reha-Krantz, L. J. (1998). Identification of a transient excision intermediate at the crossroads between DNA polymerase extension and proofreading pathways. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3507–3512. doi: 10.1073/pnas.95.7.3507
- Beckman, R. A., Mildvan, A. S., and Loeb, L. A. (1985). On the fidelity of DNA replication: manganese mutagenesis *in vitro*. *Biochemistry* 24, 5810–5817. doi: 10.1021/bi00342a019
- Bedinger, P., Munn, M., and Alberts, B. M. (1989). Sequence-specific pausing during *in vitro* DNA replication on double-stranded DNA templates. *J. Biol. Chem.* 264, 16880–16886.
- Bessman, M. J., and Reha-Krantz, L. J. (1977). Studies on the biochemical basis of spontaneous mutation. V. Effect of temperature on mutation frequency. *J. Mol. Biol.* 116, 115–123. doi: 10.1016/0022-2836(77)90122-X
- Blanco, L., Bernad, A., and Salas, M. (1983). *Phi29 DNA Polymerase*. U.S. Patent 5,198,543.
- Bloom, L. B., Otto, M. R., Eritja, R., Reha-Krantz, L. J., Goodman, M. F., and Beechem, J. M. (1994). Pre-steady state kinetic analysis of sequence-dependent nucleotide excision by the 3'-exonuclease activity of bacteriophage T4 DNA polymerase. *Biochemistry* 33, 7576–7586. doi: 10.1021/bi00190a010
- Bourn, W., Faurholm, B., and Foskett, J. (2011). *Modified DNA Polymerases*. U.S. Patent Application US2011/0269211.
- Brutlag, D., and Kornberg, A. (1972). Enzymatic synthesis of deoxyribonucleic acid: XXXVI. A proofreading function for the 3' to 5' exonuclease activity in deoxyribonucleic acid polymerases. *J. Biol. Chem.* 247, 241–248.
- Clarke, L. A., Rebela, C. S., Goncalves, J., Boavida, M. G., and Jordan, P. (2001). PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. *J. Clin. Pathol. Mol. Pathol.* 54, 351–353. doi: 10.1136/mp.54.5.351
- Connolly, B. A. (2009). Recognition of deaminates bases by archaeal family-B DNA polymerases. *Biochem. Soc. Trans.* 37, 65–68. doi: 10.1042/BST0370065
- Devireddy, L. R., and Jones, C. (1998). Alternative splicing of the latency-related transcript of bovine herpesvirus 1 yields RNAs containing unique open reading frames. *J. Virol.* 72, 7294–7301.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986
- Elisseeva, E., Mandal, S. S., and Reha-Krantz, L. J. (1999). Mutational and pH studies of the 3'→5' exonuclease activity of bacteriophage T4 DNA polymerase. *J. Biol. Chem.* 273, 25151–25158. doi: 10.1074/jbc.273.25.25151
- Fazekas, A. J., Steeves, R., and Newmaster, S. G. (2010). Improving sequencing quality from PCR products. *Biotechniques* 48, 277–285. doi: 10.2144/000113369
- Fidalgo da Silva, E., Mandal, S. S., and Reha-Krantz, L. J. (2002). Using 2-aminopurine fluorescence to measure incorporation of incorrect nucleotides by wild type and mutant bacteriophage T4 DNA polymerases. *J. Biol. Chem.* 277, 40640–40649. doi: 10.1074/jbc.M203315200
- Fidalgo da Silva, E., and Reha-Krantz, L. J. (2000). Dinucleotide repeat expansion catalyzed by bacteriophage T4 DNA polymerase *in vitro*. *J. Biol. Chem.* 275, 431528–431535. doi: 10.1074/jbc.M004594200
- Fidalgo da Silva, E., and Reha-Krantz, L. J. (2007). DNA polymerase proofreading: active site switching catalyzed by the bacteriophage T4 DNA polymerase. *Nucleic Acids Res.* 35, 5452–5463. doi: 10.1093/nar/gkm591
- Franklin, M. C., Wang, J., and Steitz, T. A. (2001). Structure of a replicating complex of a replicating complex of a pol family DNA polymerase. *Cell* 105, 657–667. doi: 10.1016/S0092-8674(01)00367-1
- Goodman, M. F., and Reha-Krantz, L. J. (1999). Synthesis of fluorophore-labeled DNA. *US Patent* 5945312.
- Goodwin, P. M., Affleck, R. L., Ambrose, W. P., Demas, J. N., Jett, J. H., Martin, J. C., et al. (1995). Progress towards DNA sequencing at the single molecule level. *Exp. Tech. Phys.* 41, 279–294. doi: 10.1016/S0168-1656(03)00006-3
- Hacker, K. J., and Alberts, B. M. (1994). The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. *J. Biol. Chem.* 269, 24221–24228.
- Hamilton, S. C., Farchaus, J. W., and Davis, M. C. (2001). DNA polymerases as engines for biotechnology. *Biotechniques* 31, 370–383.
- Hariharan, C., Bloom, L. G., Helquist, S. A., Kool, E. T., and Reha-Krantz, L. J. (2006). Dynamics of nucleotide incorporation: snapshots revealed by 2-aminopurine fluorescence studies. *Biochemistry* 45, 2836–2844. doi: 10.1021/bi051644s
- Hariharan, C., and Reha-Krantz, L. J. (2005). Using 2-aminopurine fluorescence to detect bacteriophage T4 DNA polymerase-DNA complexes that are important

- for primer extension and proofreading reactions. *Biochemistry* 44, 15674–15684. doi: 10.1021/bi051462y
- Hogg, M., Aller, P., Konigsberg, W., Wallace, S. S., and Doublie, S. (2007). Structural and biochemical investigation of the role in proofreading of a β hairpin loop found in the exonuclease domain of a replicative DNA polymerase of the B family. *J. Biol. Chem.* 282, 1432–1444. doi: 10.1074/jbc.M605675200
- Hogg, M., Cooper, W., Reha-Krantz, L., and Wallace, S. (2006). Kinetics of error generation in homologous B-family DNA polymerases. *Nucleic Acids Res.* 34, 2528–2535. doi: 10.1093/nar/gkl300
- Hogg, M., Wallace, S. S., and Doublie, S. (2004). Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J.* 23, 1483–1493. doi: 10.1038/sj.emboj.7600150
- Huang, C.-C., Hearst, J. E., and Alberts, B. M. (1981). Two types of replication proteins increase the rate at which T4 DNA polymerase traverses the helical regions in a single-stranded DNA template. *J. Biol. Chem.* 256, 4087–4094.
- Indiani, C., McInemey, P., Georgescu, R., Goodman, M. F., and O'Donnell, M. (2005). A sliding clamp toolbelt binds high and low fidelity DNA polymerases simultaneously. *Mol. Cell* 19, 805–819. doi: 10.1016/j.molcel.2005.08.011
- Jarvis, T. C., Newport, J. W., and von Hippel, P. H. (1991). Stimulation of the processivity of the DNA polymerase of bacteriophage T4 by the polymerase accessory proteins. *J. Biol. Chem.* 266, 1830–1840.
- Jin, Y. H., Garg, P., Stitich, C. M., Al-Refaie, H., Sterling, J. F., Murray, L. J., et al. (2004). The multiple biological roles of the 3'-5' exonuclease of *Saccharomyces cerevisiae* DNA polymerase delta require switching between polymerase and exonuclease domains. *Mol. Cell. Biol.* 25, 261–471.
- Jose, D., Datta, K., Johnson, N. P., and von Hippel, P. H. (2009). Spectroscopic studies of position-specific DNA “breathing” fluctuations at replication forks and primer-template junctions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4231–4236. doi: 10.1073/pnas.0900803106
- Joyce, C. M. (2004). T4 replication: what does “processivity” really mean? *Proc. Natl. Acad. Sci. U.S.A.* 101, 8255–8256. doi: 10.1073/pnas.0402850101
- Kunkel, T. A., Loeb, L. A., and Goodman, M. F. (1984). On the fidelity of DNA replication. The accuracy of T4 DNA polymerase in copying X174 DNA *in vitro*. *J. Biol. Chem.* 259, 1539–1545.
- Kunkel, T. A., Patel, S. S., and Johnson, K. A. (1994). Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6830–6834. doi: 10.1073/pnas.91.15.6830
- Li, L. (2004). A Novel DNA Polymerase Delta Mutant of *Saccharomyces cerevisiae* with Sensitivity to Phosphonoacetic Acid, Ph.D. thesis, University of Alberta, Alberta, Canada.
- Li, L., Murphy, K. M., Kanevets, U., and Reha-Krantz, L. J. (2005). Sensitivity of phosphonoacetic acid: a new phenotype to probe DNA polymerase (*in Saccharomyces cerevisiae*). *Genetics* 170, 569–580. doi: 10.1534/genetics.104.040295
- Li, V., Hogg, M., and Reha-Krantz, L. J. (2010). Identification of a new motif in family B DNA polymerases by mutational analyses of the bacteriophage T4 DNA polymerase. *J. Mol. Biol.* 400, 295–308. doi: 10.1016/j.jmb.2010.05.030
- Lin, T.-C., Rush, J., Spicer, E. K., and Konigsberg, W. H. (1987). Cloning and expression of T4 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 84, 7000–7004. doi: 10.1073/pnas.84.20.7000
- Mandal, S. S., Fidalgo da Silva, E., and Reha-Krantz, L. J. (2002). Using 2-aminopurine fluorescence to detect base unstacking in the template strand during nucleotide incorporation by the bacteriophage T4 DNA polymerase. *Biochemistry* 41, 4399–4406. doi: 10.1021/bi015723p
- Marchand, B., Tchesnokov, E. P., and Götte, M. (2007). The pyrophosphate analogue foscarnet traps the pre-translocational state of HIV-1 reverse transcriptase in a brownian ratchet model of polymerase translocation. *J. Biol. Chem.* 282, 3337–3346. doi: 10.1074/jbc.M607710200
- Marquez, L. A., and Reha-Krantz, L. J. (1996). Using 2-aminopurine fluorescence and mutational analysis to demonstrate an active role of bacteriophage T4 DNA polymerase in strand separation required for 3'→5' exonuclease activity. *J. Biol. Chem.* 271, 28903–28911. doi: 10.1074/jbc.271.46.28903
- Mizrahi, V., Benlovic, P., and Benkovic, S. J. (1986). Mechanism of DNA polymerase I: exonuclease/polymerase activity switch and DNA sequence dependence of pyrophosphorolysis and misincorporation reactions. *Proc. Natl. Acad. Sci. U.S.A.* 83, 5769–5773. doi: 10.1073/pnas.83.16.5769
- Moldovan, G.-L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. *Cell* 129, 665–679. doi: 10.1016/j.cell.2007.05.003
- Murphy, K., Darmawan, H., Schultz, A., Fidalgo da Silva, E., and Reha-Krantz, L. J. (2006). A method to select for mutator DNA polymerase deltas in *Saccharomyces cerevisiae*. *Genome* 49, 403–410. doi: 10.1139/G05-106
- Muzyczka, N., Poland, R. L., and Bessman, M. J. (1972). Studies on the biochemical basis of spontaneous mutation. I. A comparison of the deoxyribonucleic acid polymerases of mutator, antimutator, and wild type strains of bacteriophage T4. *J. Biol. Chem.* 247, 7116–7122.
- Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. J., and Kunkel, T. A. (2008). Division of labor at the eukaryotic replication fork. *Cell* 30, 137–144. doi: 10.1016/j.molcel.2008.02.022
- Niimi, A., Limsirichaikul, S., Yoshida, S., Iwai, S., Masutani, C., Hanaoka, F., et al. (2004). Palm mutants in DNA polymerases alpha and eta alter DNA replication fidelity and translesion activity. *Mol. Cell. Biol.* 24, 2734–2746. doi: 10.1128/MCB.24.7.2734-2746.2004
- Pagès, V., and Fuchs, R. P. P. (2002). How DNA lesions are turned into mutations within cells. *Oncogene* 21, 8957–8966. doi: 10.1038/sj.onc.1206006
- Patel, P. H., Kawate, H., Adman, E., Ashback, M., and Loeb, L. A. (2001). A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J. Biol. Chem.* 276, 5044–5051. doi: 10.1074/jbc.M008701200
- Patel, P. H., and Loeb, L. A. (2000). DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5095–5100. doi: 10.1073/pnas.97.10.5095
- Pursell, A. E., Isoz, I., Lundström, E. B., Johansson, E., and Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 317, 127–130. doi: 10.1126/science.1144067
- Rao, B. S. (1994). Pausing of simian virus 40 DNA replication fork movement *in vivo* by (dG-dA)_n-(dT-dC)_n tracts. *Gene* 140, 233–237. doi: 10.1016/0378-1119(94)90549-5
- Reddy, M. K., Weitzel, S. E., and von Hippel, P. H. (1992). Processive proofreading is intrinsic to T4 DNA polymerase. *J. Biol. Chem.* 267, 14157–14166.
- Reha-Krantz, L. J. (1988). Amino acid changes coded by bacteriophage T4 DNA polymerase mutator mutants. Relating structure to function. *J. Mol. Biol.* 202, 711–724. doi: 10.1016/0022-2836(88)90552-9
- Reha-Krantz, L. J. (1995). Learning about DNA polymerase function by studying antimutator DNA polymerases. *Trends Biochem. Sci.* 20, 136–140. doi: 10.1016/S0968-0004(00)88987-2
- Reha-Krantz, L. J. (1998). Regulation of DNA polymerase exonucleolytic proofreading activity: studies of bacteriophage T4 “antimutator” DNA polymerases. *Genetics* 148, 1551–1557.
- Reha-Krantz, L. J. (2008). Recent patents of gene sequences relative to DNA polymerases. *Recent Pat. DNA Gene Seq.* 2, 145–163. doi: 10.2174/187221508786241729
- Reha-Krantz, L. J. (2009). “Chapter 21: the use of 2-aminopurine fluorescence to study DNA polymerase function,” in *Molecular Methods in DNA Replication*, eds S. Vengrova and J. S. Dalggaard (New York: Humana Press), 381–396.
- Reha-Krantz, L. J. (2010). DNA polymerase proofreading: multiple roles maintain genome stability. *Biochem. Biophys. Acta* 1804, 1049–1063. doi: 10.1016/j.bbapap.2009.06.012
- Reha-Krantz, L. J., Hariharan, C., Subuddhi, U., Xia, S., Zhao, C., Beckman, J., et al. (2011). Structure of the 2-aminopurine–cytosine base pair formed in the polymerase active site of the RB69 Y567A-DNA. *Biochemistry* 50, 10136–10149. doi: 10.1021/bi2014618
- Reha-Krantz, L. J., Liesner, E. M., Parmaksizoglu, S., and Stocki, S. (1986). Isolation of bacteriophage T4 DNA polymerase mutator mutants. *J. Mol. Biol.* 189, 261–272. doi: 10.1016/0022-2836(86)90508-5
- Reha-Krantz, L. J., Marquez, L. A., Eliseeva, E., Baker, R. P., Bloom, L. B., Dunford, B. H., et al. (1998). The proofreading pathway of bacteriophage T4 DNA polymerase. *J. Biol. Chem.* 273, 22969–22976. doi: 10.1074/jbc.273.36.22969
- Reha-Krantz, L. J., and Nonay, R. L. (1993). Genetic and biochemical studies of bacteriophage T4 DNA polymerase 3'→5' exonuclease activity. *J. Biol. Chem.* 268, 27100–27108.

- Reha-Krantz, L. J., and Nonay, R. L. (1994). Motif A of bacteriophage T4 DNA polymerase: role in primer extension and DNA replication fidelity. Isolation of new antimutator and mutator DNA polymerases. *J. Biol. Chem.* 269, 5635–5643.
- Reha-Krantz, L. J., Nonay, R. L., and Stocki, S. (1993). Bacteriophage T4 DNA polymerase mutations that confer sensitivity to the PP_i analog phosphonoacetic acid. *J. Virol.* 67, 60–66.
- Reha-Krantz, L. J., and Wong, C. (1996). Selection of bacteriophage T4 antimutator DNA polymerases: a link between proofreading and sensitivity to phosphonoacetic acid. *Mutat. Res.* 350, 9–16. doi: 10.1016/0027-5107(95)00085-2
- Ridlington, J. W., and Butler, L. G. (1972). Yeast inorganic pyrophosphatase. I. Binding of pyrophosphate, metal ion, and metal ion-pryophosphate complexes. *J. Biol. Chem.* 247, 7303–7307.
- Roth, A. C., Nossal, N. G., and Englund, P. T. (1982). Rapid hydrolysis of deoxynucleoside triphosphates accompanies DNA synthesis by T4 DNA polymerase and T4 accessory proteins 44/62 and 45. *J. Biol. Chem.* 257, 1267–1273.
- Rush, J., Lin, T.-C., Quinones, M., Spicer, E. K., Douglas, I., Williams, K., et al. (1989). The 44P subunit of the T4 DNA polymerase accessory protein complex catalyzes ATP hydrolysis. *J. Biol. Chem.* 264, 10943–10953.
- Sakamoto, A. N., Stone, J. E., Kissling, G. E., McCulloch, S. D., Pavlov, Y. I., and Kunkel, T. A. (2007). Mutator alleles of yeast DNA polymerase zeta. *DNA Repair* 6, 1829–1838. doi: 10.1016/j.dnarep.2007.07.002
- Shamoo, Y., Adari, H., Konigsberg, W. H., Williams, K. R., and Chase, J. R. (1986). Cloning of T4 gene 32 and expression of the wild-type protein under lambda promoter PL regulation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8844–8848. doi: 10.1073/pnas.83.23.8844
- Spacciapoli, P., and Nossal, N. G. (1994). A single mutation in bacteriophage T4 DNA polymerase (A727V, tsL141) decreases its processivity as a polymerase and increases its processivity as a 3'→5' exonuclease. *J. Biol. Chem.* 269, 438–446.
- Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D., et al. (1988). Primary structure of the T4 DNA polymerase. Evolutionary relatedness to eucaryotic and other prokaryotic DNA polymerases. *J. Biol. Chem.* 263, 7478–7486.
- Stocki, S. A., Nonay, R. L., and Reha-Krantz, L. J. (1995). Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 3'→5' exonuclease activities. *J. Mol. Biol.* 254, 15–28. doi: 10.1006/jmbi.1995.0595
- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., et al. (1966). Frameshift mutations and the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* 31, 77–84. doi: 10.1101/SQB.1966.031.01.014
- Streisinger, G., and Owen, J. (1985). Mechanism of spontaneous, and induced frameshift mutation in bacteriophage T4. *Genetics* 109, 633–659.
- Subuddhi, U., Hogg, M., and Reha-Krantz, L. J. (2008). Use of 2-aminopurine fluorescence to study the role of the beta hairpin in the proofreading pathway catalyzed by the phage T4 and RB69 DNA polymerases. *Biochemistry* 47, 6130–6137. doi: 10.1021/bi800211f
- Tabor, S., and Richardson, C. S. (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci. U.S.A.* 86, 4076–4080. doi: 10.1073/pnas.86.11.4076
- Tabor, S., and Richardson, C. S. (1990). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Effect of pyrophosphorylation and metal ions. *J. Biol. Chem.* 265, 8322–8328.
- Tsujikawa, L., Weinfield, M., and Reha-Krantz, L. J. (2003). Differences in replication of a DNA template containing an ethyl phosphotriester by T4 DNA polymerase and *Escherichia coli* DNA polymerase I. *Nucleic Acids Res.* 31, 4965–4972. doi: 10.1093/nar/gkg722
- Tullius, M. V., Harth, G., and Horwitz, M. A. (2001). High extracellular levels of *Mycobacterium tuberculosis* glutamine synthetase and superoxide dismutase in actively growing cultures are due to high expression and extracellular stability rather than to a protein-specific export mechanism. *Infect. Immun.* 69, 6348–6363. doi: 10.1128/IAI.69.10.6348–6363.2001
- Venkatesan, R. N., Hsu, J. J., Lawrence, N. A., Preston, B. D., and Loeb, L. A. (2006). Mutator phenotypes caused by a substitution at a conserved motif A residue in eukaryotic DNA polymerase. *J. Biol. Chem.* 281, 4486–4494. doi: 10.1074/jbc.M510245200
- Venkatesan, R. N., Treuting, P. M., Fuller, E. D., Goldsby, R. E., and Norwood, T. H., et al. (2007). Mutation at the polymerase active site in mouse DNA polymerase increases genomic instability and accelerates tumorigenesis. *Mol. Cell. Biol.* 27, 7669–7682. doi: 10.1128/MCB.00002-07
- Wang, F. J., and Ripley, S. L. (1998). The spectrum of acridine resistant mutants of bacteriophage T4 reveals cryptic effects of the tsL141 DNA polymerase allele on spontaneous mutagenesis. *Genetics* 148, 1655–1665.
- Wang, M., Xia, S., Blaha, G., Steitz, T. A., Konigsberg, W., and Wang, J. (2011). Insights into base selectivity from the 1.8 Å resolution structure of an RB69 DNA polymerase ternary complex. *Biochemistry* 50, 581–590. doi: 10.1021/bi101192f
- Werner, J. H., Cai, H., Jett, J. H., Reha-Krantz, L. J., Keller, R. A., and Keller, R. A. (2003). Progress towards single-molecule DNA sequencing: a one color demonstration. *J. Biotechnol.* 102, 1–14. doi: 10.1016/S0168-1656(03)00006-3
- Xie, S., Wood, M., Bradley, M. J., De la Cruz, E. M., and Konigsberg, W. H. (2013). Alteration in the cavity size adjacent to the active site of the RB69 DNA polymerase changes its conformational dynamics. *Nucleic Acids Res.* 41, 9077–9089. doi: 10.1093/nar/gkt674
- Yang, J., Zhuang, Z., Roccasecca, R. M., Trakselis, M. A., and Benkovic, S. J. (2004). The dynamic processivity of the T4 DNA polymerase during replication. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8289–8294. doi: 10.1073/pnas.0402625101
- Zahn, K. E., Tchesnokov, E. P., Götte, M., and Doublie, S. (2011). Phosphonoformic acid inhibits viral replication by trapping the closed form of the DNA polymerase. *J. Biol. Chem.* 286, 25246–25255. doi: 10.1074/jbc.M111.248864
- Zhong, X., Pedersen, L. C., and Kunkel, T. A. (2008). Characterization of a replicative DNA polymerase mutant with reduced fidelity and increased translesion synthesis capacity. *Nucleic Acids Res.* 36, 3892–3904. doi: 10.1093/nar/gkn312
- Zhu, B. (2014). Bacteriophage T7 DNA polymerase – sequenase. *Front. Microbiol.* 5:181. doi: 10.3389/fmicb.2014.00181

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 June 2014; accepted: 07 July 2014; published online: 04 August 2014.

*Citation: Reha-Krantz LJ, Woodgate S and Goodman MF (2014) Engineering processive DNA polymerases with maximum benefit at minimum cost. *Front. Microbiol.* 5:380. doi: 10.3389/fmicb.2014.00380*

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Reha-Krantz, Woodgate and Goodman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present

Cheng-Yao Chen*

Protein Engineering Group, Illumina, San Diego, CA, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Suleyman Yildirim, Walter Reed Army Institute of Research, USA

Andreas Marx, University of Konstanz, Germany

***Correspondence:**

Cheng-Yao Chen, Protein Engineering Group, Illumina, 5200 Illumina Way, San Diego, CA 92122, USA
e-mail: cchen2@illumina.com

Next-generation sequencing (NGS) technologies have revolutionized modern biological and biomedical research. The engines responsible for this innovation are DNA polymerases; they catalyze the biochemical reaction for deriving template sequence information. In fact, DNA polymerase has been a cornerstone of DNA sequencing from the very beginning. *Escherichia coli* DNA polymerase I proteolytic (Klenow) fragment was originally utilized in Sanger's dideoxy chain-terminating DNA sequencing chemistry. From these humble beginnings followed an explosion of organism-specific, genome sequence information accessible via public database. Family A/B DNA polymerases from mesophilic/thermophilic bacteria/archaea were modified and tested in today's standard capillary electrophoresis (CE) and NGS sequencing platforms. These enzymes were selected for their efficient incorporation of bulky dye-terminator and reversible dye-terminator nucleotides respectively. Third generation, real-time single molecule sequencing platform requires slightly different enzyme properties. Enterobacterial phage ϕ 29 DNA polymerase copies long stretches of DNA and possesses a unique capability to efficiently incorporate terminal phosphate-labeled nucleoside polyphosphates. Furthermore, ϕ 29 enzyme has also been utilized in emerging DNA sequencing technologies including nanopore-, and protein-transistor-based sequencing. DNA polymerase is, and will continue to be, a crucial component of sequencing technologies.

Keywords: Sanger sequencing, chain terminators, reversible terminators, sequencing-by-synthesis, DNA polymerase, next-generation sequencing, protein engineering

INTRODUCTION

Since the advent of enzymatic dideoxy-DNA sequencing by Fredric Sanger (Sanger et al., 1977), sequencing DNA/RNA has become standard practice in most molecular biology research. The proliferation of next-generation sequencing (NGS) technologies has further transformed modern biological and biomedical research. Today, large-scale whole genome sequencing has become routine in life science research. Although technical advances in current NGS technologies have dramatically changed the way nucleic acids are sequenced, the engine ultimately responsible for these modern innovations remains unchanged. Like Sanger sequencing, today's NGS technologies, with the exception of oligonucleotide-based ligation sequencing (Drmanac et al., 2010), still require a DNA polymerase to carry out the necessary biochemical reaction for replicating template sequence information. This unique, polymerase-dependent sequencing approach is generally referred to as DNA sequencing-by-synthesis (SBS), because the consecutive sequencing reaction concurrently generates a newly synthesized DNA strand as a result.

However, unlike Sanger sequencing, DNA polymerases utilized in NGS technologies are more diverse and tailor-made. The Klenow enzyme, a proteolytic fragment of *Escherichia coli* DNA polymerase I, was originally utilized in Sanger's dideoxy chain-terminating DNA sequencing chemistry (Sanger et al., 1977). This enzyme was chosen for its efficient incorporation of 2', 3'-dideoxynucleotides (ddNTPs) that leads to chain termination of

DNA synthesis (Atkinson et al., 1969). From this humble beginning, followed by a robust sequencing chemistry improvement, the nucleotide substrates used for DNA sequencing became larger and bulkier. First, four fluorescent dyes with distinct, non-overlapping optical spectra were attached to either purine or pyrimidine bases, respectively, and even the terminal gamma phosphate of four (A, T, C, and G) nucleotides for the ease of signal detection (Smith et al., 1986; Ju et al., 2006; Guo et al., 2008, 2010; Eid et al., 2009; Korlach et al., 2010). Second, the 3' hydroxyl group on deoxyribose of four nucleotides was replaced with a larger, cleavable chemical group used to reversibly terminate DNA synthesis (Ju et al., 2006; Guo et al., 2008, 2010). As a result, the original Klenow enzyme no longer efficiently incorporated these newly modified nucleotides. DNA polymerases with different enzymatic properties were required for improving the nucleotide incorporation reactions. Fortunately, the adoption of NGS sequencing in life science research allowed a rapid expansion of organism-specific, genome sequence information accessible via public database. Various DNA polymerases from mesophilic/thermophilic viruses, bacteria, and archaea were discovered and later screened for efficient incorporation of modified nucleotides in new DNA sequencing methods. A pool of new, advantageous DNA polymerases from a wide variety of microorganisms were selected and served as protein backbones for further improvement via protein engineering or directed enzyme evolution (Patel and Loeb, 2001).

Evolved DNA polymerases with improved biochemical performances were ultimately utilized for each, unique sequencing technology.

This article briefly covers (1) the progression of decades' enzymatic DNA sequencing methods reliant on functions of DNA polymerase for synthesizing new DNA strands; (2) the novel properties of DNA polymerase that are required for high-precision DNA sequencing; (3) the influence of nucleotide modifications on DNA polymerase research that ultimately lead to improved sequencing performance; (4) the application of DNA polymerases in emerging DNA sequencing methods. Readers interest in learning more about other sequencing methods can refer to these literatures (Landegren et al., 1988; Ding et al., 2012) for more information.

OVERVIEW OF DNA POLYMERASE FAMILIES AND FUNCTIONS

Since the discovery of DNA polymerase I in *E. coli* by Arthur Kornberg's group in the late 1950s (Lehman et al., 1958a,b), multiple DNA polymerases have been discovered, purified and characterized from bacteria, eukaryotes, archaea, and their viruses. The expansion of organism-specific, genome sequence information accessible via public database, together with advanced search-algorithms based on DNA polymerase structure–function relationships, have reduced the time necessary for identification of additional, putative DNA polymerases from a variety of sources (Burgers et al., 2001). Based on the phylogenetic relationships of *E. coli* and human DNA polymerases, DNA polymerases are generally classified into seven main families: A (*E. coli* Pol I), B (*E. coli* Pol II), C (*E. coli* Pol III), D, X (human Pol β -like), Y (*E. coli* Pol IV and V and TLS polymerases), and RT (reverse transcriptase) (Burgers et al., 2001; Langhorst et al., 2012). All living organisms, except viruses, harbor multiple types of DNA polymerases for

cellular functions. Interestingly, neither bacteria, eukaryotes nor archaea contain all families of DNA polymerases. As summarized in Table 1, the family C DNA polymerases are unique to bacteria, and have not been found in either eukaryotes or archaea (Hübscher et al., 2010; Langhorst et al., 2012). Likewise, the family D polymerases are restricted to archaea (Euryarchaeota), and do not exist in bacteria or eukaryotes (Hübscher et al., 2010; Langhorst et al., 2012). Another characteristic exclusive to archaeal DNA polymerases is the presence of intervening sequences (inteins) within the polymerase coding genes (Perler et al., 1992). These inteins cause in-frame insertions in archaeal DNA pols and must be spliced out in order to form mature enzymes (Hodges et al., 1992).

The basic function of DNA polymerases (cellular DNA replicases) are to faithfully replicate the organism's whole genome and pass down the correct genetic information to future generations. In bacteria, family C DNA polymerases, such as Pol III holoenzyme in *E. coli* or *Bacillus subtilis*, are the key element for driving chromosomal replication and thus absolutely mandatory for cell viability (Gefter et al., 1971; Nusslein et al., 1971; Gass and Cozzarelli, 1973, 1974). Besides the Pol III holoenzyme, the A-family Pol I also participates in bacterial DNA replication (Olivera and Bonhoeffer, 1974). Pol I contains a separate 5' to 3' exonuclease, independent of the DNA polymerase domain, that can remove RNA primers and concurrently fill in the nucleotide gaps between Okazaki fragments during lagging strand DNA synthesis (Okazaki et al., 1971; Konrad and Lehman, 1974; Xu et al., 1997). Unlike bacterial cells, eukaryotic B-family DNA polymerases, such as Pol δ and ϵ in human and yeast, are responsible for nuclear chromosomal replication (Miyabe et al., 2011). Recent studies in yeast by Thomas Kunkel's group suggest that Pol δ and ϵ divide their roles during DNA replication and are responsible for lagging and leading strand DNA synthesis, respectively (Pursell et al., 2007; Kunkel and

Table 1 | Families and properties of cellular DNA replicases (Kunkel, 2004; Hübscher et al., 2010; Greenough et al., 2014).

Polymerase family	Bacteria (<i>E. coli</i>)	Eukaryotes (human)	Archaea	Viruses	3' to 5' exonuclease activity	**Error rate (fidelity)	Enzymes used in assays
A	Pol I [<i>pol A</i>]	Pol γ (p140/p55/p55) Pol θ (p100/p90/p80) Pol ν	N.A.	T3, T5, T7	Yes	$\sim 10^{-5}$ – 10^{-7}	Klenow, KlenTaq, Taq, Bst, Bsu, T7
B	Pol II [<i>pol B</i>]	Pol α /primase (p180/p68/Pri2/Pri1) Pol δ (p125/p66/p50/p12) Pol ϵ (p260/p59/p17/p12) Pol ζ (p350/p24)	Pol BI (p180/p68/Pri2/Pri1) Pol BII Pol BIII	HSV-1, RB69, T4, ϕ 29	Yes	$\sim 10^{-6}$	T4, ϕ 29, 9 $^{\circ}$ N, KOD1, Pfu, Vent
C	Pol III [<i>pol C</i>] core ($\alpha/\epsilon/\theta$)	N.A.	N.A.	N.A.	Yes	$\sim 10^{-6}$	N.A.
D	*N.A.	N.A.	Pol D (DP2/DP1)	N.A.	Yes	10^{-4} – 10^{-5}	N.A.

In the bacterial column, the gene for each corresponding protein is indicated in the bracket. In the eukaryotic and archaeal column, the components of each holoenzyme are listed in the parentheses. *N.A. denotes "not applicable." **The unit for "Error Rate" is one error per incorporated base.

Burgers, 2008; Nick McElhinny et al., 2008; Miyabe et al., 2011). In archaea, both B- and D-family pols are involved in genomic replication. However, the role of each Pol *in vivo* remains controversial. From biochemical studies, both Pol B and D enzymes from hyperthermophilic *Pyrococcus abyssi* are proposed to function together in DNA replication (Henneke et al., 2005). In contrast, a recent genetic study in *Thermococcus kodakarensis* showed that Pol D alone is sufficient for cell viability and genomic replication which argues that Pol D, rather than Pol B, is the main replicative DNA polymerase in this archaeon (Cubanova et al., 2013). It is possible that the requirements for Pol B and D enzymes in DNA replication are different in separate phyla of Archaea.

In summary, all DNA polymerases engaged in cellular genome replication, regardless of origin, have the following common features (See **Table 1**): (1) they appear to form a multi-subunit enzyme complex (holoenzyme); (2) they all possess an intrinsic 3' to 5' exonuclease, or proofreading activity, that removes misincorporated nucleotides immediately after nucleotide incorporation to ensure high-fidelity of DNA synthesis (**Figure 3A**). In contrast to the major cellular DNA polymerases, functions of X, Y, and RT families of Pols are more diverse and specialized in many DNA processes, such as DNA repair, translesion synthesis, and eukaryotic telomere maintenance (Hübscher et al., 2010). None of these Pols have any intrinsic 3' to 5' proofreading exonuclease activity and are thus more error-prone during DNA synthesis (Kunkel and Bebenek, 2000; Kunkel, 2004, 2009).

CHOOSING THE RIGHT DNA POLYMERASE FOR DNA SEQUENCING

Growing numbers of DNA polymerases, each with distinct functions, provide abundant enzymatic resources for improving current and emerging DNA sequencing techniques. However, not all families of DNA polymerases are suitable for high-precision DNA sequencing reactions. To be considered, and ultimately applied for a particular method of sequencing, the DNA polymerase should possess the following properties:

- (1) The pol has to be a DNA-dependent DNA polymerase. Some X and RT-family enzymes do not require a DNA template for replication and are thus not suitable for DNA sequencing; for instance, X-family terminal deoxynucleotidyl transferases (Tdt) are template-independent DNA polymerases which catalyze the addition of deoxynucleotides (dNTPs) to the 3'-OH ends of DNA in the absence of a DNA template (Kato et al., 1967; Coleman et al., 1974). Similarly, RT-family eukaryotic telomerasess are ribonucleoproteins which utilize their own, endogenous RNA template for elongation at the telomeric DNA ends (Morin, 1989; Blackburn et al., 2006). These enzymes bypass the requirement of a DNA template to function and cannot be used for DNA sequencing.
- (2) The pol should rapidly incorporate nucleotides. Despite the diverse functions among DNA polymerases, the basic mechanism of nucleotide incorporation remains relatively fixed. All replicative DNA pols require a duplex primer-template DNA with a free 3'-OH group at the primer terminus, all four nucleoside triphosphates (dATP, dTTP, dCTP, and dGTP),

and catalytic, divalent cations (Mg^{2+} or Mn^{2+} , etc.) for the sequencing reaction. Addition of nucleotides to the 3' end of a primer by DNA pols proceeds through a highly ordered, temporal mechanism. The minimal catalytic mechanism of single-nucleotide incorporation by DNA pol has been proposed (Donlin et al., 1991; Johnson, 1993) and is illustrated in **Figure 1**. A brief description for each reaction step can be found in the figure legend. As shown in **Figure 1**, the nucleotide incorporation efficiency (specificity) of a DNA polymerase ($k_{pol}/k_{d,dNTP}$) is determined by the rate of phosphodiester bond formation (k_{pol}) and the binding constant for the cognate nucleotide ($k_{d,dNTP}$; Wong et al., 1991; Johnson, 1993). DNA pols with a faster nucleotide incorporation rate and lower $k_{d,dNTP}$ (large k_{pol} and small $k_{d,dNTP}$) can catalyze DNA synthesis much more efficiently. In this aspect, none of the X and Y-family pols can meet this requirement. Both X and Y-family Pols have much lower nucleotide incorporation efficiency (Brown et al., 2011a,b) compared to cellular DNA replicases from A, B, C, or D-family enzymes (Patel et al., 1991; Wong et al., 1991; Bloom et al., 1997; Zhang et al., 2009). Therefore, they are not ideal for DNA sequencing.

- (3) The pol must have high replicative fidelity to minimize systematic sequencing errors. In order to accurately read DNA template sequence information, the DNA pol must faithfully

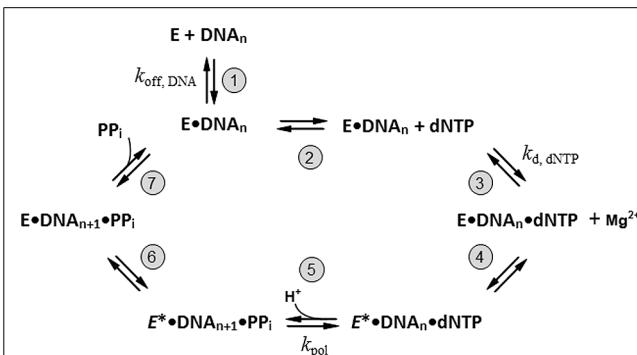


FIGURE 1 |The minimal catalytic steps required for single-nucleotide incorporation by DNA polymerase. The addition of nucleotide to the 3' end of a primer by DNA polymerase passes through a temporally ordered mechanism. The reaction begins with the binding of free DNA polymerase (E) to a duplex primer/template DNA complex (DNA_n) resulting in a binary enzyme–DNA complex ($E \bullet DNA_n$; step 1). The $k_{off, DNA}$ represents the rate of enzyme dissociation from the $E \bullet DNA_n$ complex. Addition of the correct nucleotide (dNTP) in the presence of divalent cations, such as Mg^{2+} , promotes the enzyme–DNA-dNTP ternary complex formation ($E \bullet DNA_n \bullet dNTP$; step 2 and 3). The $k_{d, dNTP}$ denotes the nucleotide binding constant of the enzyme. The binding of the dNTP induces the first conformational change of the enzyme in the ternary complex ($E^* \bullet DNA_n \bullet dNTP$; step 4; Wong et al., 1991). The actual chemistry happens (step 5). The phosphodiester bond is formed between the α -phosphate of the incoming dNTP and 3'-OH of the primer terminus (DNA_{n+1}). The chemical reaction generates a pyrophosphate (PP_i) and proton molecule (H^+). This is followed by a second conformational change of the enzyme (step 6), which allows the final release of the PP_i leaving group (step 7). The nucleotide incorporation cycle is complete after PP_i release. If the enzyme remains associated with DNA, a new round of nucleotide addition will continue until the enzyme dissociates from the DNA (processive synthesis).

incorporate the correct, matched nucleotides along the DNA template. The fidelity of nucleotide incorporation by X, Y, and RT Pols range from $\sim 10^{-1}$ to 10^{-4} error per base incorporated, two to three orders of magnitude lower than high-fidelity cellular DNA polymerases from A, B, or C-family enzymes (Kunkel, 2004). These repair pols generally make errors during DNA synthesis (Kunkel and Bebenek, 2000; Kunkel, 2004, 2009) and are not appropriate for high-precision DNA sequencing applications.

- (4) The pol should possess long, intrinsic, replicative processivity. The processivity of DNA polymerase is defined as the number of dNTPs incorporated during complex formation with a primer/template (P/T) DNA. As illustrated in **Figure 1**, the processivity of DNA pol is directly related to two parameters: (1) the rate of dNTP incorporation by the enzyme (k_{pol} of step 5); (2) the enzyme's dissociation rate from the enzyme–DNA binary complex ($k_{off,DNA}$ of step 1). Under these parameters, the enzyme remains associated with the template DNA, it carries out sequential rounds of nucleotide incorporation until it dissociates from the binary complex (**Figure 1**, steps 2–7). Theoretically, processivity of the DNA polymerase can be estimated by calculating the ratio of $k_{pol}/k_{off,DNA}$. Amongst DNA polymerases, only viral DNA polymerases have unusually intrinsic, long processivity. For instance, the enterobacterial phage ϕ 29 DNA polymerase (a B-family enzyme) possesses intrinsic, long, replicative processivity and can replicate its own genomic DNA ($\sim 18,000$ base pairs) *in vitro* without any accessory cofactors (Blanco and Salas, 1985). In contrast, most cellular DNA replicases from A, B, C, and D families are distributive, and limited to only a few nucleotide incorporations. These DNA replicases must physically interact with their processivity factors, including β -sliding clamp in bacteria, and PCNA in eukaryotes and archaea, in order to achieve a long processivity during DNA replication (Johnson and O'Donnell, 2005). No X, Y, or RT-family enzymes are processive.
- (5) The pol should function as a monomer for ease of protein production and further modification. As illustrated in **Table 1**, most A, B, C, and D-family DNA replicases form a multi-subunit enzyme complex (holoenzyme). Components of these replicative holoenzymes are difficult to purify, and whole enzyme complexes are very challenging to reconstitute. Therefore, these types of enzyme complexes are seldom used in any DNA sequencing chemistry.

In summary, to fulfill the above requirements for high-precision DNA sequencing, only A-family enzymes from bacteria and phage viruses (such as T5 and T7 phages), and B-family pols from bacterial viruses (such as T4, Rb69, and ϕ 29 phages), bacteria, and archaea (*Vent*, $9^{\circ}N$, *Pfu*, and *KOD1*) have been evaluated for sequencing chemistry development (See **Table 1**). All family A and B enzymes have an associated, intrinsic 3' to 5' exonuclease proofreading activity. When these enzymes incorporate an incorrect nucleotide at the primer terminus, the enzymes' ability to extend the primer terminus diminishes, and allows the nascent DNA strand to migrate to the 3' exonuclease site for excision (See **Figure 3A**; Donlin et al., 1991; Joyce

and Steitz, 1994; Patel and Loeb, 2001). This unique partitioning mechanism of the 3' exonuclease proofreading domain among A and B-family polymerases is disfavored for DNA sequencing. It causes asynchronous DNA sequencing reactions and generates systematic sequencing errors (**Figures 3B,C**). Therefore, the majority of A and B-family pols used for DNA sequencing are either lacking, or have an attenuated, 3' exonuclease proofreading activity.

NUCLEOTIDE SUBSTRATES FOR THE GENERATIONS OF DNA POLYMERASE-BASED SEQUENCING

Generations of DNA polymerase-based sequencing methods and their corresponding commercial platforms are summarized in **Table 2**. As shown in **Table 2**, all methods require a DNA polymerase to catalyze the necessary biochemical reaction for extracting DNA sequence information. The fundamental difference amongst these technologies is the type of nucleotide substrate incorporated. The structures of these nucleotides are illustrated in **Figure 2**. More in-depth information regarding these nucleotides can be found in the following articles (Metzker et al., 1996; Lee et al., 1997; Kumar et al., 2005; Metzker, 2010; Chen et al., 2013a). From classical Sanger sequencing to modern NGS technologies, the nucleotide substrates used for sequencing have changed over time. In the original Sanger sequencing method, four 2', 3'-ddNTPs (**Figure 2B**) are utilized (Sanger et al., 1977). Unlike normal dNTPs (**Figure 2A**), the ddNTPs lack the 3'-hydroxyl group (3'-OH), which is required for the phosphodiester bond formation between the incorporating nucleotide and primer terminus. Once ddNTPs are incorporated by the DNA polymerase, they terminate further addition of nucleotides from the primer terminus, and cease elongation of the DNA chain (Atkinson et al., 1969). Besides the utilization of ddNTPs, Sanger's protocol requires a set of radioisotope-labeled primers in four, separate (A, T, C, and G) reactions. The resulting dideoxy-terminated DNA fragments must be analyzed side-by-side using slab gel electrophoresis while sequence information is deduced via autoradiography (Sanger et al., 1977). The procedure itself is extremely time consuming and further compounded by low data output. This makes such an approach insufficient at meeting the growing demand for high-throughput DNA sequencing.

To simplify and subsequently automate Sanger's method, Leroy Hood's group, then at California Institute of Technology, invented the first fluorescent sequencing (dye-primer) method based on Sanger's approach (Smith et al., 1986). In Hood's revised protocol, the primers used for sequencing reactions are covalently attached to four distinct colors of fluorophores at the 5'-end, corresponding to each of the A, T, C, and G reactions in Sanger sequencing. The advantages to this approach are (1) the four reaction mixtures can be combined and analyzed in a single sequencing lane; (2) the results can be directly monitored by a computer-aided fluorescence detection system, specifically matched to the emission spectra of the four dyes. These advantages allow DNA sequence information to be analyzed automatically by the computer.

Hood's dye-primer method simplifies traditional Sanger sequencing processes but it is not, however, completely ideal

Table 2 | Generations of DNA polymerase-based DNA sequencing technologies.

Institution or company	Instrumentation	Sequencing methods	Nucleotide substrates	Detection from DNA polymerase reaction
MRC	Sanger sequencing	DNA chain termination and fragment analysis by gel	2', 3'-dideoxynucleotides	Nucleotide incorporation
Applied biosystems/life technologies	ABI genetic analyzer series	DNA chain termination and fragment analysis by CE	Dye-terminators	Nucleotide incorporation
Illumina	GA/MiSeq/HiSeq	Stepwise SBS	Reversible dye-terminators	Nucleotide incorporation
Qiagen/IBS	Max-Seq/Mini-20			
*Helicos biosciences	HeliScope	Stepwise single-molecule SBS	3'-OH unblocked reversible dye-terminators	Nucleotide incorporation
Pacific biosciences	PACBIO RS II	Real-time single-molecule SBS	γ -phosphate-labeled nucleotides	Nucleotide incorporation
Roche/454 life sciences	GS FLX/GS Junior	Sequential SBS	Nature dNTPs	PP _i release
Ion torrent/life technologies	Ion PGM/proton	Sequential SBS	Nature dNTPs	H ⁺ release

*On November 15, 2012, Helicos Biosciences filed for Chap. 11 bankruptcy.

for fully automated DNA sequencing, mainly due to the four, separate reactions still required. To solve this problem, the fluorescently labeled chain-terminating ddNTPs (dye-terminators) were soon introduced by Prober et al. (1987) from DuPont. Similar to the dye-primers, a set of fluorescently distinguished fluorophores are covalently attached to each of four ddNTPs (See **Figure 2C**). Adaptation of dye-terminators for Sanger sequencing workflow makes the four, base-specific chain termination reactions happen in one, single reaction tube. DNA polymerase is able to simultaneously incorporate four dye-terminators and generate the terminated DNA pieces for sequence analysis (Rosenthal and Charnock-Jones, 1992, 1993). The speed and throughput of dye-terminator sequencing was drastically improved when the automated capillary-array electrophoresis (CAE) was adopted for DNA analysis (Drossman et al., 1990; Luckey et al., 1990; Zagursky and McCormick, 1990; Dovichi and Zhang, 2000).

The dye-terminator-CE method has greatly improved sequencing performance and has become the laboratory standard for DNA sequencing over the past few decades. However, the technique itself is still very limited, especially for large-scale, whole genome sequencing. Increasing the sequencing throughput of dye-terminator-CE chemistry requires additional capillary tubes to be implemented. This becomes impractical for the application of high-throughput, multiplexing sequencing that is capable of sequencing millions of different DNA strands concurrently. To alleviate this limitation, reversible dye-terminators were introduced to the modified, dye-terminating sequencing scheme. Similar to dye-terminators (**Figure 2C**), reversible dye-terminators (**Figure 2D**) are also missing the 3'-OH group needed for DNA polymerase extension of the primer terminus. Incorporation of these modified nucleotides by DNA polymerase terminates DNA chain elongation (Bentley et al., 2008; Guo et al., 2008; Hutter et al., 2010). When these reversible dye-terminators are used in parallel with immobilization of DNA molecules on a solid-state surface,

the individual DNA sequence can be directly ascertained from the base-specific, terminated DNA molecules recognized by the fluorescent imaging system (Bentley et al., 2008; Guo et al., 2008, 2010). As a result, the requirements for capillary electrophoresis (CE) analysis in a typical dye-terminator approach are no longer necessary, and millions of different DNA molecules can be sequenced simultaneously. Differentiating themselves from dye-terminators, reversible dye-terminators contain cleavable chemical groups at the 3' position of the pentose and linker region, located between the base and attached fluorophore (**Figure 2D**; Bentley et al., 2008; Guo et al., 2008; Hutter et al., 2010). These cleavable chemical groups can be removed in order to restore the normal 3'-OH group of deoxyribose and maintain the integrity of bases attached with dye. DNA chains can thus be further extended by the DNA polymerase and incorporation can resume once more in the next reaction cycle (Bentley et al., 2008; Guo et al., 2008, 2010). A similar sequencing scheme was also carried out using another class of reversible dye-terminators with normal 3'-OH groups (Wu et al., 2007; Pushkarev et al., 2009; Litosh et al., 2011; Gardner et al., 2012). These 3' unblocked, reversible terminators possess both chemical blockage group and fluorescent dye attached to the same base (**Figure 2E**), and can be removed by either chemical cleavage or UV light (Pushkarev et al., 2009; Litosh et al., 2011).

In both classes of reversible dye-terminators, cleavage of the linker group carrying the fluorescent dye leaves extra chemical molecules on the normal purine and pyrimidine bases. These molecular remnants may perturb the protein-DNA interaction and eventually impact the sequencing performance of the DNA polymerase (Metzker, 2010; Chen et al., 2013a). To circumvent this concern, terminal γ -phosphate, fluorescently labeled nucleoside polyphosphates (**Figure 2F**) were developed for the more advanced, third-generation DNA sequencing technique (Kumar et al., 2005; Korlach et al., 2010). There are two major

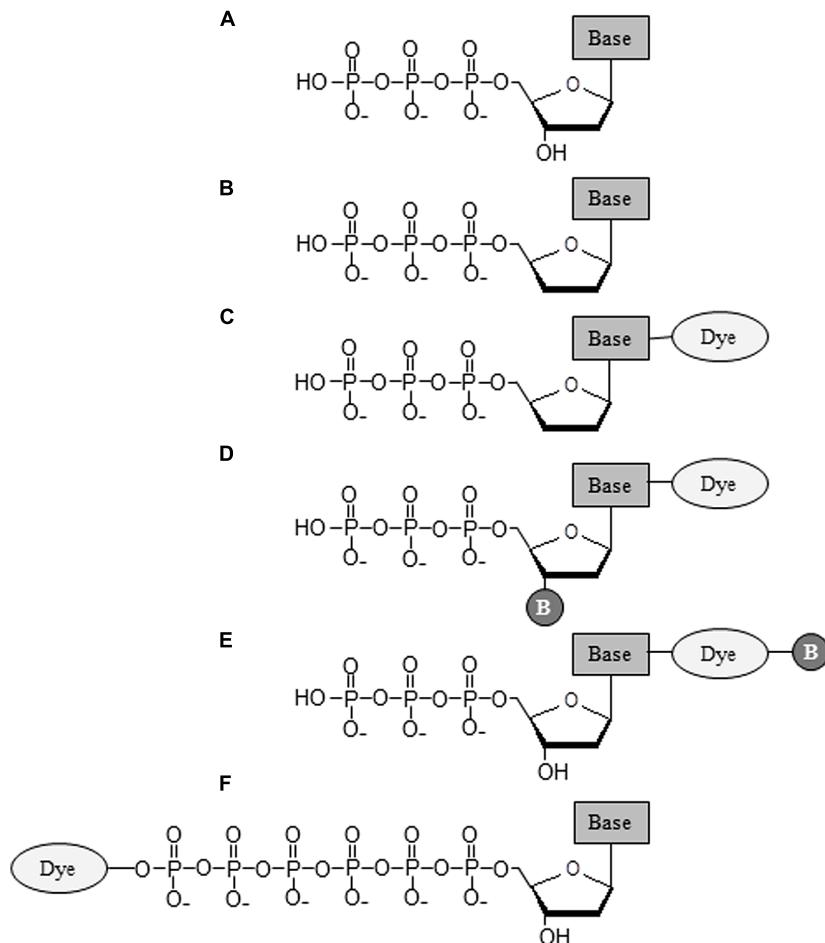


FIGURE 2 | Structures of nucleotides utilized in the generations of DNA polymerase-based sequencing methods. (A) Deoxynucleotides (dNTPs); **(B)** 2', 3'-dideoxynucleotides (ddNTPs); **(C)** Dye-terminators; **(D)** Reversible

sible dye-terminators; **(E)** 3'-OH unblocked reversible dye-terminators; **(F)** Dye-labeled hexaphosphate nucleotides. The “Base” in the diagram represents an A, T, C or G base, and “B” indicates a cleavable chemical blockage group.

advantages of performing DNA sequencing with γ -phosphate-labeled nucleotides over conventional chain terminators. First, the nucleotides, once incorporated, don't generate a molecular scar on the newly synthesized DNA, and second, they enable real-time, single-molecule SBS (Korlach et al., 2010). Because the phosphoryl transfer reaction only occurs between the 3'-OH group of the primer terminus and α -phosphate of the incoming nucleotide, the conclusion of each enzymatic reaction results in one nucleotide addition to the primer terminus plus a pyrophosphate (PP_i) leaving group (Figure 1, steps 5–7; Steitz, 1997, 1999). Hence, any fluorophore covalently attached to the PP_i leaving group will be released after nucleotide addition to the primer terminus, and thus leave no molecular vestige in the DNA. Since the added nucleotide possesses no blockage group to hinder DNA elongation from the primer terminus, the sequencing reaction can continue uninterrupted.

Finally, there are no DNA scar issues for both pyrosequencing technology (Ronaghi et al., 1996, 1998), which detects the release of PP_i after nucleotide addition by DNA polymerase, and

semiconductor-based proton sequencing technique (Rothberg et al., 2011), which monitors the proton (H⁺) release during phosphodiester bond formation between the 3'-OH and α -phosphate of incoming nucleotide. Both technologies utilize natural nucleoside triphosphates (dNTPs) for their sequencing reactions (Table 2 and Figure 2A).

CHALLENGES OF RAPIDLY EVOLVING NUCLEOTIDE SUBSTRATES ON DNA POLYMERASE RESEARCH

A series of nucleotide modifications, created for rapidly changing DNA polymerase-based sequencing technologies has created a daunting task for DNA polymerase researchers to look for, design or evolve compatible enzymes for ever-changing DNA sequencing chemistries. From the beginning, A-family *E. coli* DNA polymerase I (Pol I) or its proteolytic (Klenow) fragment was chosen by Dr. Sanger for his dideoxy-sequencing chemistry (Sanger et al., 1977). This was the only DNA polymerase available at the time and, quite fortunately, tolerated incorporation of 2', 3'-ddNTPs (Atkinson et al., 1969). However, Pol I effectively discriminates between a deoxy- and dideoxyribose in the

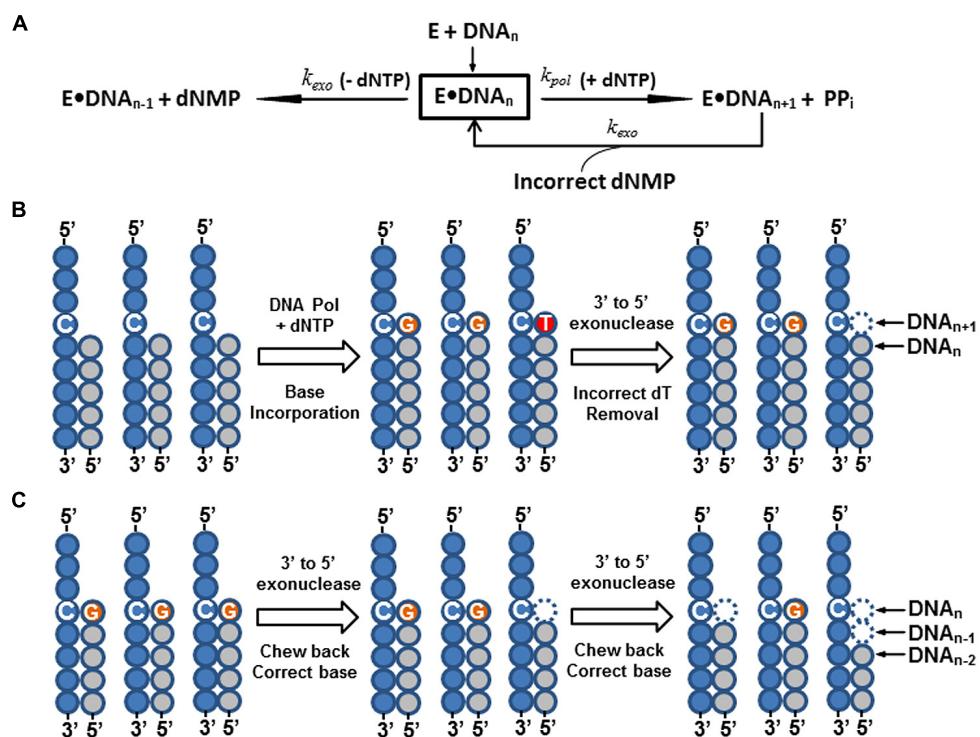


FIGURE 3 | Intrinsic 3' to 5' exonuclease activity of DNA polymerase and its impact on DNA sequencing reactions. **(A)** A simplified kinetic model illustrating the proofreading function and nucleotide excision activity of 3' to 5' exonuclease of DNA polymerase (Donlin et al., 1991; Johnson, 1993). As shown in the figure, when a free DNA polymerase (E) is mixed with a duplex primer/template DNA complex (DNA_n), they form a stable, binary enzyme–DNA complex ($\text{E} \bullet \text{DNA}_n$). In the presence of nucleotide (+dNTP) and divalent cations (Mg^{2+} , Mn^{2+} , etc.), the enzyme rapidly incorporates (k_{pol}) a single-nucleotide base to the primer terminus (DNA_{n+1}) and concurrently drives release of free pyrophosphate (PP_i). However, when an incorrect nucleotide is misincorporated by the enzyme, it causes a base-pair mismatch at the primer terminus (DNA_{n+1} ; Panel **B**, middle cartoon, a dC:dT mismatch). This mismatched nucleotide base at the primer terminus greatly impedes the DNA polymerase's capability to incorporate the next nucleotide base (greatly reduced k_{pol} value) and triggers a rapid transfer of DNA primer strand to the intrinsic 3' to 5' exonuclease domain. The mismatched nucleotide base is then removed (incorrect deoxynucleoside monophosphate, dNMP) by the 3' to 5'

exonuclease (k_{exo}). Once the mismatched nucleotide base is excised by the 3' to 5' exonuclease, the corrected primer strand is shifted back to the DNA polymerase catalytic domain ($\text{E} \bullet \text{DNA}_n$). As a result, the misincorporated nucleotide is removed and the enzyme is ready to incorporate the correct nucleotide (see Panel **B**, left to right cartoons). In addition to the base-mismatched proofreading function of the 3' to 5' exonuclease domain, it will also gradually chew back the primer strand (DNA_{n-1} , DNA_{n-2} , etc.) and release dNMPs in the absence of nucleotide (-dNTP; see Panel **C**, left to right cartoons). An asynchronous DNA sequencing reaction occurs when the sequencing DNA polymerase misincorporates a nucleotide base (Panel **B**), or the DNA sequencing primer is chewed back by the enzyme's 3' to 5' exonuclease (Panel **C**). The outcome of both reactions produces a non-uniform duplex primer-template DNA for DNA sequencing (Panels **B,C**, the right cartoons), and causes systematic DNA sequencing errors. In the panels **B,C**, each filled circle indicates a nucleotide base. A string of filled-gray circles represents the primer strand, and a string of filled-blue circles is the template DNA strand. Specific bases (dC, dG, and dT) are indicated inside the circles.

nucleoside triphosphate, and does not incorporate ddNTPs very well (Atkinson et al., 1969). In fact, the incorporation rate of ddNTP by Pol I is several hundred-fold slower than that of normal dNTPs and is also sequence context-dependent (Tabor and Richardson, 1989). This sequence-specific ddNTP incorporation by Pol I creates non-uniform band intensities on the sequencing gel. This phenomenon becomes increasingly problematic, especially in the dye-primer/terminator sequencing, because the method of sequence information retrieval relies on the interpretation of fluorescent intensity of each dideoxy-terminated DNA band from the gel or capillary tubes. Similar results were reported with thermostable, Family A, *Thermus aquaticus* (*Taq*) DNA polymerase I (Innis et al., 1988).

In contrast, phage T7 DNA polymerase does not distinguish dNTPs from ddNTPs, and incorporates both types of nucleotides

at nearly equal efficiencies (Tabor and Richardson, 1987; Brändis et al., 1996). Thus, the intensities of dideoxy-terminated bands are significantly more uniform with T7 pol in Sanger sequencing. To understand the molecular basis for this discrepancy, sequence analysis and biochemical studies were conducted among these three, A-family enzymes. The results indicate that a single phenylalanine to tyrosine residue change (Y526) on T7 pol, homologous position (F672), of a highly conserved finger motif (motif B) in A-family pols greatly reduces the enzyme's ability to select against ddNTPs (Tabor and Richardson, 1995). Biochemical studies further confirm that mutant Pol I, or *Taq*, carrying a F672Y or F667Y mutation, respectively, loses its discriminatory ability for ddNTPs, and thus incorporates ddNTPs very efficiently (Patel and Loeb, 2001). Additionally, these two mutant proteins were demonstrated to incorporate fluorescein-

and rhodamine-labeled dye-terminators, three orders of magnitude more efficiently than their wild-type parent enzymes (Tabor and Richardson, 1995). Subsequently, T7, F672Y Pol I, and F667Y *Taq* pols were all used for manual and automated Sanger sequencing (Tabor and Richardson, 1987, 1989; Rosenthal and Charnock-Jones, 1992; Tabor and Richardson, 1995). However, *Taq* pol has become preferred for dye-terminator sequencing, because the enzyme has several advantages over Pol I or T7. The enzyme is more readily purified and modifiable for further improvement. It also has no intrinsic, 3' to 5' exonuclease proofreading activity, and is active over a broad range of temperatures (Innis et al., 1988). The thermostability of *Taq* pol became essential for sequencing after the PCR-based “cycle sequencing” approach was introduced (Rosenthal and Charnock-Jones, 1993).

The Phe to Tyr mutation at position 667 on conserved motif B of *Taq* pol only addresses the deoxy- and dideoxyribose selectivity problem in dye-terminator sequencing. The enzyme, like Pol I, possesses bias. Uneven ddNTP incorporation results in variable DNA band intensities, and unequal peak heights in CE analysis, creating unwanted sequencing errors (Parker et al., 1995; Li et al., 1999). Kinetic analysis reveals that *Taq* pol favors ddGTP incorporation over other ddNTPs, with a much more robust nucleotide incorporation rate (k_{pol} ; Brandis et al., 1996). To investigate the cause of ddGTP bias, structural analysis of all four, ddNTP-trapped ternary complexes of the large fragment of *Taq* pol (Klentaq1) was implemented. The data reveals a selective interaction between the guanidinium side chain of arginine residue 660 (R660) and the O6/N7 atoms of the guanine base of the incoming ddGTP. Substitution of the Arg660 residue with a negatively charged aspartic acid completely eliminates preference for ddGTP incorporation. The R660D/F667Y double mutant of *Taq* pol greatly improves dye-terminator sequencing quality and accuracy (Li et al., 1999).

Although the F667Y mutation on *Taq* pol greatly improves the enzyme's incorporation efficiency for dideoxy-dye-terminators, the improvement becomes marginal for the reversible dye-terminators, which carry larger chemical blocking groups than the normal 3'-OH at the 3' position of deoxyribose (Bentley et al., 2008; Guo et al., 2008; Chen et al., 2010, 2013a; Hutter et al., 2010). The 3' reversible terminating group is normally linked to the deoxyribose of the nucleotide through the oxygen atom of 3'-OH. A series of 3'-O-blocking groups have been developed including 3'-O-allyl (Ruparel et al., 2005; Wu et al., 2007), 3'-O-(2-nitrobenzyl) (Wu et al., 2007), and 3'-O-azidomethylene (Bentley et al., 2008). Serendipitously, reversible dye-terminators bearing either blockage group were found to be incorporated well by a variant of archaeal 9°N DNA polymerase (a B-family Pol) of hyperthermophilic *Thermococcus* sp. 9°N-7 (Southworth et al., 1996; Ruparel et al., 2005; Ju et al., 2006; Bentley et al., 2008). The enzyme variant bearing A485L and Y409V double mutations on conserved motifs A and B, respectively, of the DNA polymerase shows enhanced preference for incorporating both acyclic and dideoxy dye-terminators over the parent enzyme (Gardner and Jack, 2002). The same mutational effects were also found in enzyme mutants possessing homologous mutations in other archaeal, B-family DNA polymerase

species (Gardner and Jack, 1999; Gardner et al., 2004). Similarly, the analogous combination of mutations (P410L/A485T) at the same conserved protein regions of closely related, B-family DNA polymerase *Thermococcus* sp. JDF-3 also shows an additive effect on improving dye-terminator incorporation (Arezi et al., 2002). Furthermore, an A485L variant of 9°N DNA pol, termed Therminator DNA polymerase commercially, was recently demonstrated to efficiently incorporate 3'-OH unblocked dye-terminators with a terminating 2-nitrobenzyl moiety attached to hydroxymethylated nucleobases (Gardner et al., 2012). Thus, mutations at these two conserved protein motifs of archaeal, B-family DNA polymerase might affect the enzyme's selectivity and tolerance for modifications and substitutions on the deoxyribose and nucleobase.

Recently, a more rational approach was taken to search for variants of *Taq* pol that can accept new types of reversible terminators possessing a 3'-ONH₂ blocking group (dNTP-ONH₂; Chen et al., 2010). Using the structure-guided reconstruction of ancestral DNA sequence analysis on *Taq* pol, a library of 93 protein variants carrying different combinations of mutations were designed and screened for the ability to incorporate dNTP-ONH₂ in primer-extension assays. One beneficial mutation (L616A) on *Taq* pol was identified. The L616A *Taq* enzyme variants incorporated both dNTP-ONH₂ and ddNTPs faithfully and efficiently.

The path toward acquisition of a compatible DNA polymerase for incorporation of fluorescent, terminal polyphosphate-labeled nucleotides has not been so straightforward. Historically, the specificities of DNA polymerases toward γ-phosphate modified dNTPs are found to be very different, due to the various degrees of steric effects of substituted chemical groups on each enzyme's dNTP binding pocket (Arzumanov et al., 1996; Martynov et al., 1997). For instance, a bulky 2, 4-dinitrophenyl group substitution at the γ-phosphate of dNTP is a good substrate for the RT-family AMV RT, but is not acceptable for A or B-family DNA polymerases (Alexandrova et al., 1998). Similar findings were reported with the bis-(2'-deoxynucleoside) 5', 5'-triphosphates (Victorova et al., 1999). HIV-RT utilizes this type of γ-phosphate modified nucleotide very effectively, while *E. coli* Pol I and *Taq* pol do not. Interestingly, in the same study, both Pol I and *Taq* pol were found to incorporate the bis-(2'-deoxynucleoside) 5', 5'-tetraphosphates more efficiently than the triphosphate analog (Victorova et al., 1999). Thus, the addition of an extra-phosphate moiety to the terminal γ-phosphate of dNTP seems to attenuate the steric effects on the enzyme. Alternatively stated, the extra phosphate spacer, linked to the terminal γ-phosphate of dNTP, makes the modified nucleotide better tolerated by the enzyme. Indeed, when nucleotide incorporation rates were evaluated with fluorescent, terminal phosphate-labeled nucleoside polyphosphates containing 3, or more, phosphates at the 5'-position of the nucleoside, the nucleotides possessing greater than three phosphates were more effective substrates for A and B-family DNA polymerases (Kumar et al., 2005). Later studies proved both dye-labeled nucleoside penta/hexaphosphates (dN5Ps and dN6P) alone can be used by enterobacterial phage φ29 DNA polymerase for incorporating thousands of bases in length, approaching natural dNTP rates

(Korlach et al., 2008, 2010). This unique, long, replicative processivity of φ29 DNA pol, together with intrinsic, superior capability of incorporating dye-labeled, terminal polyphosphate nucleotides plays a key role in real-time, single-molecule SBS (Korlach et al., 2010).

APPLICATIONS OF DNA POLYMERASE FOR EMERGING SEQUENCING TECHNOLOGIES

In contrast to current, SBS approaches, emergent DNA sequencing methods rely on unconventional applications of DNA polymerase. These techniques utilize DNA polymerase as a traditional incorporating enzyme, and alternatively as a molecular motor, responsible for controlled DNA translocation across the protein nanopore. Traditional, nanopore-based, SBS uses commercial Therminator γ DNA polymerase, a variant⁹N DNA pol, to incorporate terminal, γ-phosphate-labeled nucleoside tetraphosphates. These modified nucleotides are coupled with four, different-length PEG-coumarin tags corresponding to base A, T, C, and G (Kumar et al., 2012). DNA sequence information can be ascertained by measuring current (*amp*) fluctuations of the orderly, released PEG-coumarin tags through the α-hemolysin nanopore following DNA polymerase incorporation. A related, but fundamentally different approach involves mutant *Mycobacterium smegmatis* porin A (MspA) nanopore, φ29 DNA polymerase, and natural dNTPs (Manrao et al., 2012). In this approach, the enzyme functions as both DNA replicative enzyme, and molecular motor, which control the speed of DNA translocation through the MspA nanopore.

Besides the nanopore-based sequencing approach, a protein, transistor-based sequencing method, leveraging electrical conductance measurement of φ29 DNA polymerase reactions has been reported (Chen et al., 2013b). Unfortunately, this study is currently called into question, and the merits of this particular method must be reevaluated (Chen et al., 2013b).

CONCLUSION

Since the introduction of the first enzymatic DNA sequencing by Frederic Sanger in the mid-1970s, decades of scientific research on various DNA polymerases, starting with Arthur Kornberg's enzyme discovery in the mid-1950s, have provided the basic understanding of how these enzymes function and replicate DNAs, further cementing the foundation for improving enzyme properties and applications in current, and future, DNA polymerase-based sequencing technologies. The large-scale of organism-specific, genome research reveals the intrinsic diversity and unique characteristics of DNA polymerases present in all kingdoms of life, including their viruses. Diverse DNA polymerases with distinct functions and properties provide a large pool of natural protein variants that can be tested, and later utilized, for continuously evolving sequencing-chemistries. Tailor-made protein variants designed via protein engineering or directed-enzyme evolution have created powerful protein-engines that have propelled the progression of DNA sequencing technologies over the past few decades. Without a doubt, DNA polymerase has been, and will continue to remain, a crucial component of future sequencing technologies.

ACKNOWLEDGMENTS

The author thanks Ali Nikoomanzar for editing the manuscript, and Dr. Lawrence A. Loeb, Eddie Fox, and Thomas Lie for critical reading the manuscript.

REFERENCES

- Alexandrova, L. A., Skoblov, A. Y., Jasko, M. V., Victorova, L. S., and Krayevsky, A. A. (1998). 2'-Deoxynucleoside 5'-triphosphates modified at alpha-, beta- and gamma-phosphates as substrates for DNA polymerases. *Nucleic Acids Res.* 26, 778–786. doi: 10.1093/nar/26.3.778
- Arezi, B., Hansen, C. J., and Hogrefe, H. H. (2002). Efficient and high fidelity incorporation of dye-terminators by a novel archaeal DNA polymerase mutant. *J. Mol. Biol.* 322, 719–729. doi: 10.1016/S0022-2836(02)00843-4
- Arzumanov, A. A., Semizarov, D. G., Victorova, L. S., Dyatkina, N. B., and Krayevsky, A. A. (1996). Gamma-phosphate-substituted 2'-deoxynucleoside 5'-triphosphates as substrates for DNA polymerases. *J. Biol. Chem.* 271, 24389–24394. doi: 10.1074/jbc.271.40.24389
- Atkinson, M. R., Deutscher, M. P., Kornberg, A., Russell, A. F., and Moffatt, J. G. (1969). Enzymatic synthesis of deoxyribonucleic acid. XXXIV. Termination of chain growth by a 2',3'-dideoxyribonucleotide. *Biochemistry* 8, 4897–4904. doi: 10.1021/bi00840a037
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. doi: 10.1038/nature07517
- Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006). Telomeres and telomerase: the path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nat. Med.* 12, 1133–1138. doi: 10.1038/nm1006-1133
- Blanco, L., and Salas, M. (1985). Replication of phage phi 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length phi 29 DNA. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6404–6408. doi: 10.1073/pnas.82.19.6404
- Bloom, L. B., Chen, X., Fygenson, D. K., Turner, J., O'Donnell, M., and Goodman, M. F. (1997). 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.* 272, 27919–27930. doi: 10.1074/jbc.272.44.27919
- Brandis, J. W., Edwards, S. G., and Johnson, K. A. (1996). Slow rate of phosphodiester bond formation accounts for the strong bias that Taq DNA polymerase shows against 2',3'-dideoxynucleotide terminators. *Biochemistry* 35, 2189–2200. doi: 10.1021/bi951682j
- Brown, J. A., Pack, L. R., Fowler, J. D., and Suo, Z. (2011a). Pre-steady-state kinetic analysis of the incorporation of anti-HIV nucleotide analogs catalyzed by human X- and Y-family DNA polymerases. *Antimicrob. Agents Chemother.* 55, 276–283. doi: 10.1128/AAC.01229-10
- Brown, J. A., Pack, L. R., Sanman, L. E., and Suo, Z. (2011b). Efficiency and fidelity of human DNA polymerases lambda and beta during gap-filling DNA synthesis. *DNA Repair (Amst.)* 10, 24–33. doi: 10.1016/j.dnarep.2010.09.005
- Burgers, P. M., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., et al. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.* 276, 43487–43490. doi: 10.1074/jbc.R100056200
- Chen, F., Dong, M., Ge, M., Zhu, L., Ren, L., Liu, G., et al. (2013a). The history and advances of reversible terminators used in new generations of sequencing technology. *Genomics Proteomics Bioinformatics* 11, 34–40. doi: 10.1016/j.gpb.2013.01.003
- Chen, Y. S., Lee, C. H., Hung, M. Y., Pan, H. A., Chiou, J. C., and Huang, G. S. (2013b). DNA sequencing using electrical conductance measurements of a DNA polymerase. *Nat. Nanotechnol.* 8, 452–458. doi: 10.1038/nnano.2013.71
- Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., et al. (2010). Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1948–1953. doi: 10.1073/pnas.0908463107
- Coleman, M. S., Hutton, J. J., and Bollum, F. J. (1974). Terminal deoxynucleotidyl transferase and DNA polymerase in classes of cells from rat thymus. *Biochem. Biophys. Res. Commun.* 58, 1104–1109. doi: 10.1016/S0006-291X(74)80257-3
- Cubanova, L., Richardson, T., Burkhardt, B. W., Kelman, Z., Connolly, B. A., Reeve, J. N., et al. (2013). Archaeal DNA polymerase D but not DNA polymerase B is

- required for genome replication in *Thermococcus kodakarensis*. *J. Bacteriol.* 195, 2322–2328. doi: 10.1128/JB.02037-12
- Ding, F., Manosas, M., Spiering, M. M., Benkovic, S. J., Bensimon, D., Allemand, J. F., et al. (2012). Single-molecule mechanical identification and sequencing. *Nat. Methods* 9, 367–372. doi: 10.1038/nmeth.1925
- Donlin, M. J., Patel, S. S., and Johnson, K. A. (1991). Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* 30, 538–546. doi: 10.1021/bi0216a031
- Dovichi, N. J., and Zhang, J. (2000). How Capillary Electrophoresis Sequenced the Human Genome This Essay is based on a lecture given at the Analytica 2000 conference in Munich (Germany) on the occasion of the Heinrich-Emanuel-Merck Prize presentation. *Angew. Chem. Int. Ed. Engl.* 39, 4463–4468. doi: 10.1002/1521-3773(20001215)39:24<4463::AID-ANIE4463>3.0.CO;2-8
- Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327, 78–81. doi: 10.1126/science.1181498
- Drossman, H., Luckey, J. A., Kostichka, A. J., D'Cunha, J., and Smith, L. M. (1990). High-speed separations of DNA sequencing reactions by capillary electrophoresis. *Anal. Chem.* 62, 900–903. doi: 10.1021/ac00208a003
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986
- Gardner, A. F., and Jack, W. E. (1999). Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* 27, 2545–2553. doi: 10.1093/nar/27.12.2545
- Gardner, A. F., and Jack, W. E. (2002). Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases. *Nucleic Acids Res.* 30, 605–613. doi: 10.1093/nar/30.2.605
- Gardner, A. F., Joyce, C. M., and Jack, W. E. (2004). Comparative kinetics of nucleotide analog incorporation by vent DNA polymerase. *J. Biol. Chem.* 279, 11834–11842. doi: 10.1074/jbc.M308286200
- Gardner, A. F., Wang, J., Wu, W., Karouby, J., Li, H., Stupi, B. P., et al. (2012). Rapid incorporation kinetics and improved fidelity of a novel class of 3'-OH unblocked reversible terminators. *Nucleic Acids Res.* 40, 7404–7415. doi: 10.1093/nar/gks330
- Gass, K. B., and Cozzarelli, N. R. (1973). Further genetic and enzymological characterization of the three *Bacillus subtilis* deoxyribonucleic acid polymerases. *J. Biol. Chem.* 248, 7688–7700.
- Gass, K. B., and Cozzarelli, N. R. (1974). *Bacillus subtilis* DNA polymerases. *Methods Enzymol.* 29, 27–38. doi: 10.1016/0076-6879(74)29006-2
- Geftter, M. L., Hirota, Y., Kornberg, T., Wechsler, J. A., and Barnoux, C. (1971). Analysis of DNA polymerases II and III in mutants of *Escherichia coli* thermosensitive for DNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 68, 3150–3153. doi: 10.1073/pnas.68.12.3150
- Greenough, L., Menin, J. F., Desai, N. S., Kelman, Z., and Gardner, A. F. (2014). Characterization of Family D DNA polymerase from *Thermococcus* sp. 9 degrees N. *Extremophiles* doi: 10.1007/s00792-014-0646-9 [Epub ahead of print].
- Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Kim, D. H., et al. (2008). Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9145–9150. doi: 10.1073/pnas.0804023105
- Guo, J., Yu, L., Turro, N. J., and Ju, J. (2010). An integrated system for DNA sequencing by synthesis using novel nucleotide analogues. *Acc. Chem. Res.* 43, 551–563. doi: 10.1021/ar900255c
- Henneke, G., Flament, D., Hübscher, U., Querellou, J., and Raffin, J. P. (2005). The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J. Mol. Biol.* 350, 53–64. doi: 10.1016/j.jmb.2005.04.042
- Hodges, R. A., Perler, F. B., Noren, C. J., and Jack, W. E. (1992). Protein splicing removes intervening sequences in an archaea DNA polymerase. *Nucleic Acids Res.* 20, 6153–6157. doi: 10.1093/nar/20.23.6153
- Hübscher, U., Spadari, S., Villani, G., and Giovanni, M. (2010). *DNA Polymerases: Discovery, Characterization and Functions in Cellular DNA Transactions*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Hutter, D., Kim, M. J., Karalkar, N., Leal, N. A., Chen, F., Guggenheim, E., et al. (2010). Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups. *Nucleosides Nucleotides Nucleic Acids* 29, 879–895. doi: 10.1080/15257770.2010.536191
- Innis, M. A., Myambo, K. B., Gelfand, D. H., and Brow, M. A. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. U.S.A.* 85, 9436–9440. doi: 10.1073/pnas.85.24.9436
- Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* 74, 283–315. doi: 10.1146/annurev.biochem.73.011303.073859
- Johnson, K. A. (1993). Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* 62, 685–713. doi: 10.1146/annurev.bi.62.070193.003345
- Joyce, C. M., and Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63, 777–822. doi: 10.1146/annurev.bi.63.070194.004021
- Ju, J., Kim, D. H., Bi, L., Meng, Q., Bai, X., Li, Z., et al. (2006). Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19635–19640. doi: 10.1073/pnas.0609513103
- Kato, K. I., Goncalves, J. M., Houts, G. E., and Bollum, F. J. (1967). Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase. *J. Biol. Chem.* 242, 2780–2789.
- Konrad, E. B., and Lehman, I. R. (1974). A conditional lethal mutant of *Escherichia coli* K12 defective in the 5' leads to 3' exonuclease associated with DNA polymerase I. *Proc. Natl. Acad. Sci. U.S.A.* 71, 2048–2051. doi: 10.1073/pnas.71.5.2048
- Korlach, J., Bibillo, A., Wegener, J., Peluso, P., Pham, T. T., Park, I., et al. (2008). Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* 27, 1072–1083. doi: 10.1080/1525770802260741
- Korlach, J., Björnson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., et al. (2010). Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol.* 472, 431–455. doi: 10.1016/S0076-6879(10)72001-2
- Kumar, S., Sood, A., Wegener, J., Finn, P. J., Nampalli, S., Nelson, J. R., et al. (2005). Terminal phosphate labeled nucleotides: synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides Nucleotides Nucleic Acids* 24, 401–408. doi: 10.1081/NCN-200059823
- Kumar, S., Tao, C., Chien, M., Hellner, B., Balijepalli, A., Robertson, J. W., et al. (2012). PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci. Rep.* 2:684. doi: 10.1038/srep00684
- Kunkel, T. A. (2004). DNA replication fidelity. *J. Biol. Chem.* 279, 16895–16898. doi: 10.1074/jbc.R400006200
- Kunkel, T. A. (2009). Evolving views of DNA replication (in)fidelity. *Cold Spring Harb. Symp. Quant. Biol.* 74, 91–101. doi: 10.1101/sqb.2009.74.027
- Kunkel, T. A., and Bebenek, K. (2000). DNA replication fidelity. *Annu. Rev. Biochem.* 69, 497–529. doi: 10.1146/annurev.biochem.69.1.497
- Kunkel, T. A., and Burgers, P. M. (2008). Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol.* 18, 521–527. doi: 10.1016/j.tcb.2008.08.005
- Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988). A ligase-mediated gene detection technique. *Science* 241, 1077–1080. doi: 10.1126/science.3413476
- Langhorst, B. W., Jack, W. E., Reha-Krantz, L., and Nichols, N. M. (2012). Polbase: a repository of biochemical, genetic and structural information about DNA polymerases. *Nucleic Acids Res.* 40, D381–D387. doi: 10.1093/nar/gkr847
- Lee, L. G., Spurgeon, S. L., Heiner, C. R., Benson, S. C., Rosenblum, B. B., Menchen, S. M., et al. (1997). New energy transfer dyes for DNA sequencing. *Nucleic Acids Res.* 25, 2816–2822. doi: 10.1093/nar/25.14.2816
- Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958a). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J. Biol. Chem.* 233, 163–170.
- Lehman, I. R., Zimmerman, S. B., Adler, J., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958b). Enzymatic Synthesis of deoxyribonucleic acid. V. Chemical composition of enzymatically synthesized deoxyribonucleic acid. *Proc. Natl. Acad. Sci. U.S.A.* 44, 1191–1196. doi: 10.1073/pnas.44.12.1191
- Li, Y., Mitavox, V., and Waksman, G. (1999). Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9491–9496. doi: 10.1073/pnas.96.17.9491
- Litosh, V. A., Wu, W., Stupi, B. P., Wang, J., Morris, S. E., Hersch, M. N., et al. (2011). Improved nucleotide selectivity and termination of 3'-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. *Nucleic Acids Res.* 39:e39. doi: 10.1093/nar/gkq1293

- Luckey, J. A., Grossman, H., Kostichka, A. J., Mead, D. A., D'Cunha, J., Norris, T. B., et al. (1990). High speed DNA sequencing by capillary electrophoresis. *Nucleic Acids Res.* 18, 4417–4421. doi: 10.1093/nar/18.15.4417
- Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., et al. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* 30, 349–353. doi: 10.1038/nbt.2171
- Martynov, B. I., Shirokova, E. A., Jasko, M. V., Victorova, L. S., and Krayevsky, A. A. (1997). Effect of triphosphate modifications in 2'-deoxynucleoside 5'-triphosphates on their specificity towards various DNA polymerases. *FEBS Lett.* 410, 423–427. doi: 10.1016/S0014-5793(97)00577-2
- Metzker, M. L. (2010). Sequencing technologies – the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Metzker, M. L., Lu, J., and Gibbs, R. A. (1996). Electrophoretically uniform fluorescent dyes for automated DNA sequencing. *Science* 271, 1420–1422. doi: 10.1126/science.271.5254.1420
- Miyabe, I., Kunkel, T. A., and Carr, A. M. (2011). The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS Genet.* 7:e1002407. doi: 10.1371/journal.pgen.1002407
- Morin, G. B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* 59, 521–529. doi: 10.1016/0092-8674(89)90035-4
- Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M., and Kunkel, T. A. (2008). Division of labor at the eukaryotic replication fork. *Mol. Cell* 30, 137–144. doi: 10.1016/j.molcel.2008.02.022
- Nusslein, V., Otto, B., Bonhoeffer, F., and Schaller, H. (1971). Function of DNA polymerase 3 in DNA replication. *Nat. New Biol.* 234, 285–286. doi: 10.1038/newbio234285a0
- Okazaki, R., Arisawa, M., and Sugino, A. (1971). Slow joining of newly replicated DNA chains in DNA polymerase I-deficient *Escherichia coli* mutants. *Proc. Natl. Acad. Sci. U.S.A.* 68, 2954–2957. doi: 10.1073/pnas.68.12.2954
- Olivera, R. M., and Bonhoeffer, E. (1974). Replication of *Escherichia coli* requires DNA polymerase I. *Nature* 250, 513–514. doi: 10.1038/250513a0
- Parker, L. T., Deng, Q., Zakeri, H., Carlson, C., Nickerson, D. A., and Kwok, P. Y. (1995). Peak height variations in automated sequencing of PCR products using Taq dye-terminator chemistry. *Biotechniques* 19, 116–121.
- Patel, P. H., and Loeb, L. A. (2001). Getting a grip on how DNA polymerases function. *Nat. Struct. Biol.* 8, 656–659. doi: 10.1038/90344
- Patel, S. S., Wong, I., and Johnson, K. A. (1991). Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* 30, 511–525. doi: 10.1021/bi0216a029
- Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., et al. (1992). Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5577–5581. doi: 10.1073/pnas.89.12.5577
- Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., et al. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238, 336–341. doi: 10.1126/science.2443975
- Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E., and Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 317, 127–130. doi: 10.1126/science.1144067
- Pushkarev, D., Neff, N. F., and Quake, S. R. (2009). Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 27, 847–850. doi: 10.1038/nbt.1561
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242, 84–89. doi: 10.1006/abio.1996.0432
- Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363, 365. doi: 10.1126/science.281.5375.363
- Rosenthal, A., and Charnock-Jones, D. S. (1992). New protocols for DNA sequencing with dye terminators. *DNA Seq.* 3, 61–64.
- Rosenthal, A., and Charnock-Jones, D. S. (1993). Linear amplification sequencing with dye terminators. *Methods Mol. Biol.* 23, 281–296. doi: 10.1385/0-89603-248-5:281
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi: 10.1038/nature10242
- Ruparel, H., Bi, L., Li, Z., Bai, X., Kim, D. H., Turro, N. J., et al. (2005). Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5932–5937. doi: 10.1073/pnas.0501962102
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., et al. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674–679. doi: 10.1038/321674a0
- Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W., and Perler, F. B. (1996). Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus* sp. 9 degrees N-7 and mutations affecting 3'-5' exonuclease activity. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5281–5285. doi: 10.1073/pnas.93.11.5281
- Steitz, T. A. (1997). DNA and RNA polymerases: structural diversity and common mechanisms. *Harvey Lect.* 93, 75–93.
- Steitz, T. A. (1999). DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* 274, 17395–17398. doi: 10.1074/jbc.274.25.17395
- Tabor, S., and Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 84, 4767–4771. doi: 10.1073/pnas.84.14.4767
- Tabor, S., and Richardson, C. C. (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci. U.S.A.* 86, 4076–4080. doi: 10.1073/pnas.86.11.4076
- Tabor, S., and Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6339–6343. doi: 10.1073/pnas.92.14.6339
- Victorova, L., Sosunov, V., Skoblov, A., Shipotin, A., and Krayevsky, A. (1999). New substrates of DNA polymerases. *FEBS Lett.* 453, 6–10. doi: 10.1016/S0014-5793(99)00615-8
- Wong, I., Patel, S. S., and Johnson, K. A. (1991). An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* 30, 526–537. doi: 10.1021/bi00216a030
- Wu, J., Zhang, S., Meng, Q., Cao, H., Li, Z., Li, X., et al. (2007). 3'-O-modified nucleotides as reversible terminators for pyrosequencing. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16462–16467. doi: 10.1073/pnas.0707495104
- Xu, Y., Derbyshire, V., Ng, K., Sun, X. C., Grindley, N. D., and Joyce, C. M. (1997). Biochemical and mutational studies of the 5'-3' exonuclease of DNA polymerase I of *Escherichia coli*. *J. Mol. Biol.* 268, 284–302. doi: 10.1006/jmbi.1997.0967
- Zagursky, R. J., and McCormick, R. M. (1990). DNA sequencing separations in capillary gels on a modified commercial DNA sequencing instrument. *Biotechniques* 9, 74–79.
- Zhang, L., Brown, J. A., Newmister, S. A., and Suo, Z. (2009). Polymerization fidelity of a replicative DNA polymerase from the hyperthermophilic archaeon *Sulfolobus solfataricus* P2. *Biochemistry* 48, 7492–7501. doi: 10.1021/bi900532w

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 April 2014; **paper pending published:** 08 May 2014; **accepted:** 03 June 2014; **published online:** 24 June 2014.

Citation: Chen C-Y (2014) DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front. Microbiol.* 5:305. doi: 10.3389/fmicb.2014.00305
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides

Roberto Laos*, J. Michael Thomson and Steven A. Benner

Foundation for Applied Molecular Evolution, Gainesville, FL, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Michael Metzker, Baylor College of Medicine, USA

Piet Herdewijn, KU Leuven, Belgium

***Correspondence:**

Roberto Laos, Foundation for Applied Molecular Evolution, 720 SW 2nd Avenue, Suite 208, Gainesville, FL 32601, USA

e-mail: manuscripts@ffame.org

DNA polymerases have evolved for billions of years to accept natural nucleoside triphosphate substrates with high fidelity and to exclude closely related structures, such as the analogous ribonucleoside triphosphates. However, polymerases that can accept unnatural nucleoside triphosphates are desired for many applications in biotechnology. The focus of this review is on non-standard nucleotides that expand the genetic "alphabet." This review focuses on experiments that, by directed evolution, have created variants of DNA polymerases that are better able to accept unnatural nucleotides. In many cases, an analysis of past evolution of these polymerases (as inferred by examining multiple sequence alignments) can help explain some of the mutations delivered by directed evolution.

Keywords: DNA polymerases, non-standard nucleotides, AEGIS, directed evolution, CSR, protein engineering

INTRODUCTION

DNA polymerases are enzymes that catalyze the template-directed synthesis of DNA. Over billions of years, they have evolved to have the speed, specificity, and accuracy required for them to transmit valuable genetic information to and from living organisms with a level of infidelity just sufficient to support Darwinian evolution.

Currently, many DNA polymerases are used in polymerase chain reaction (PCR) and other procedures that involve the copying of nucleic acids. These include multiplexed PCR, nested PCR, reverse transcription PCR, and DNA sequencing. Polymerases are also used to incorporate modified nucleotides, including those that tag, report, or signal the presence of product DNA molecules. They are also now being used to copy sequences built from "artificially expanded genetic alphabets," which add new base pairs to the standard A:T and G:C pair. Together, these technologies are combined to allow nucleic acids to be amplified from complex samples, including saliva, blood, forensic traces, and fossil remains. Furthermore polymerases are supporting *in vitro* selection with expanded genetic alphabets to create receptors that bind to cancer cells (Sefah et al., 2014). Accordingly, the demand for new polymerase variants, especially those with specialized attributes, shows no sign of diminishing, despite the large number of polymerases already available.

This review focuses primarily on polymerase variants that accept nucleic acids having additional nucleotide "letters" that form additional nucleobase pairs. Such expanded genetic systems are being developed in many laboratories (Rappaport, 1988; Switzer et al., 1989; Ishikawa et al., 2000; Tae et al., 2001; Kool, 2002; Geyer et al., 2003; Henry and Romesberg, 2003; Minakawa et al., 2003; Benner, 2004; Hirao et al., 2004; Sismour and Benner, 2005). Some of these simply shuffle the hydrogen bonding groups that join base pairs within a Watson-Crick geometry, such as the artificially expanded genetic information system (AEGIS; Piccirilli et al.,

1990; Geyer et al., 2003). Others attempt to add hydrogen bonds to hold the pair together (Minakawa et al., 2006). Still others hope to dispense with hydrogen bonds entirely (Morales and Kool, 1999). Some polymerases have been modified without the use of directed evolution; however, these cases provide an insight on structure and function of polymerases.

Major advances in "next generation" sequencing, which requires the use of modified nucleotides and DNA polymerases, are considered in a separate review in this series (Chen, 2014).

DNA POLYMERASE FAMILIES

DNA polymerases have been classified into evolutionary families based on an analysis of their amino acid sequences. Initially two decades ago Braithwaite and Ito (1993) used an extensive compilation of the then-available sequences to classify polymerases into three families: A, B, and C. The family names indicated homology to the products of three genes: *polA*, *polB*, and *polC*, which encode for the three canonical polymerases from *Escherichia coli*: DNA polymerase I, DNA polymerase II and DNA polymerase III alpha subunit, respectively (Ito and Braithwaite, 1991; Braithwaite and Ito, 1993). The most studied polymerases belong to Family A (found in prokaryotes, eukaryotes and bacteriophages) and family B (found in prokaryotes, eukaryotes, archaea, and viruses). The D family groups polymerases from Archaea (Cann and Ishino, 1999). Families X and Y are involved in repair. The family X perform base excision repair and double-strand break repair by using their ability to fill short gaps (Moon et al., 2007; Yamtich and Sweasy, 2010). Some polymerases from the family X can perform polymerase activity without template (Berdis, 2014). The family Y groups eukaryotic polymerases (Ohmori et al., 2001) and these show less homology to the previously identified families. Most of family Y polymerases lack proofreading exonuclease domains and have a more open active site to accommodate base damage, presumably this allows them to bypass DNA lesions (Pryor et al., 2014). The RT family groups

the reverse transcriptases, including eukaryotic telomerases and reverse transcriptases found in viruses (Le Grice and Nowotny, 2014).

Early studies recognized that mild proteolysis of DNA polymerase I from *E. coli* produces two fragments, a large fragment that lacks the 5'-3' exonuclease activity and a small fragment that is then discarded. The large fragment, called the Klenow fragment, retains both the polymerization and proofreading activities of the native enzyme. The Klenow fragment yielded the first crystal structure of a family A polymerase, solved by Ollis et al. (1985). This crystal revealed a “right hand” shape with the active site being located at the “palm” which holds the catalytic amino acids, a “thumb” that binds double-stranded DNA and “fingers” where the incoming nucleotide binds and interacts with the template. The structure of *Thermus aquaticus* DNA polymerase and the analog of the Klenow fragment, the large fragment of *Thermus aquaticus* DNA polymerase (Klen-taq1) has also been studied by crystallography (Kim et al., 1995; Korolev et al., 1995).

FAMILY A POLYMERASES

Family A is the most studied of the seven DNA polymerase families. It includes many of the “workhorse” polymerase in classical molecular biology, including the Klenow fragments of *E. coli* and *Bacillus* DNA polymerase I, *Thermus aquaticus* DNA polymerase and the T7 RNA and DNA polymerases. It also includes the first DNA polymerase to be characterized enzymatically, DNA polymerase I from *E. coli*, in seminal work by Kornberg (1960).

Taq polymerase

With the advent of the PCR, it became clear that a polymerase stable to heating would be useful. Here, the DNA polymerase I from *Thermus aquaticus* (*Taq* polymerase) is widely used in PCR. *Thermus aquaticus* was isolated in 1976 from hot springs in Yellowstone National Park (Chien et al., 1976), where it thrives at 70° C. Since the enzyme can be activated by heating the sample and remains active with the high temperatures required to denature DNA strands (typically 94°C), it allows repeated cycles of denaturing, annealing and extension (thermocycling) without the need to add additional polymerase at each cycle. This made PCR a routine laboratory technique.

Eom et al. (1996) solved a co-crystal structure of *Taq* with blunt DNA duplex bound to the active site cleft. This structure had several features: (a) DNA is in an intermediate form between B and A forms. (b) Functionality from certain amino acid side chains hydrogen-bond to the N³ of purines and the O² of pyrimidines of specific residues in the duplex. (c) The 3' hydroxyl of the primer strand is near three carboxylate groups, delivered by amino acids Asp 785, Glu 786, and Asp610. These are considered to constitute the catalytic core of the enzyme.

As with its homolog, polymerase I from *E. coli*, *Taq* DNA polymerase can be cleaved to give an active fragment, called Klentaq. This fragment retains polymerase activity without one of its nucleic acid activities. Li et al. (1998) solved the crystal structures of two ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I (Klentaq1): (a) Klentaq with primer/template and dCTP; (b) Klentaq with primer template. These identified

two conformations of the polymerase: (i) an “open” conformation where the tip of the fingers of the hand is rotated 46° outward and presumably not actively performing the polymerase reaction and (ii) a “closed” conformation, which is “caught in the act” of incorporating a nucleotide. This was the first direct evidence in any DNA polymerase for a large conformational change as part of the catalytic cycle.

Motifs relevant to the engineering of Family A polymerases

Six motifs in the structure of *Taq* Pol are also conserved throughout Family A polymerases. These include motifs A, B, and C (Delarue et al., 1990). Motifs A and B are the most conserved. These two motifs are relevant to DNA polymerase fidelity and substrate specificity, which makes them of special interest to experimentalists seeking to improve the ability of investigators to obtain polymerases that accept unnatural substrates.

Motif A is found in the palm domain of the polymerase and includes the amino acids 605–617; in *Taq*, the sequence is (LLVALDYSQIELR). Within this motif, Asp 610 (bold D) cannot be changed without losing enzymatic activity, presumably, because it coordinates the metal that is directly responsible for catalysis. Glu 615 (bold E) can be changed to Asp without complete loss of activity. Tyr 611(bold Y) which is located in a hydrophobic pocket, can be replaced by a planar aromatic amino acid. The rest of the amino acids in motif A can be replaced by many amino acids without destroying catalytic activity (Patel and Loeb, 2000a).

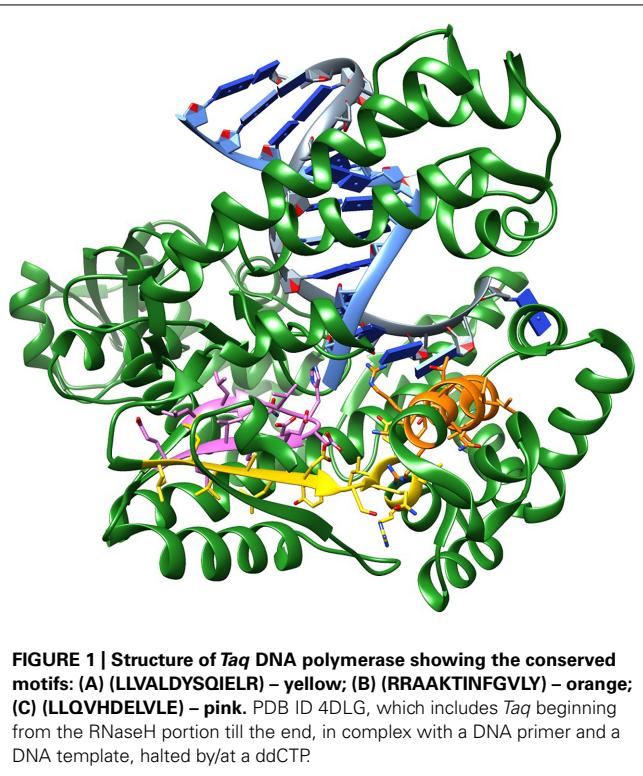
The amino acids that make up motif B are located in the fingers domain and form the O-helix, which contacts the base pair being formed in the primer extension step. This motif covers residues 659–671; in *Taq*, the sequence is (RRAAKTINFGVLY). This motif contains Arg 659 (bold R) and Lys 663 (bold K), which are known to interact with the incoming triphosphate moiety and are critical for enzymatic activity. For these reasons, they are most likely immutable. Alternatively, Phe 667 and Tyr 671 tolerate conservative substitutions as these are involved in base stacking. The remaining amino acids tolerate a wide range of substitutions. **Figure 1** shows the conserved motifs on the structure of *Taq* polymerase.

Certain substitutions within motif A and B have shown to lower fidelity without eliminating catalytic activity (Suzuki et al., 1997). This is the case of Ile614 in motif A and the Ala661Glu and Thr664Arg substitutions on motif B.

In one example, replacing an amino acid at a single site is known to change substrate specificity in a useful way. Replacement of Phe667 in *Taq* polymerase by a tyrosine eliminates the ability of the polymerase to discriminate against dideoxynucleotides. The *Taq* variant F667Y is, therefore, used for DNA sequencing. Interestingly, in T7 DNA polymerase, the replacement of Tyr526 by Phe increases the discrimination against dideoxynucleotides. This illustrates a general principle in protein engineering: rationales are best constructed *after* the replacement is made and its impact is evaluated.

FAMILY B POLYMERASES

Exploration of the natural microbiome led to the discovery of organisms that could grow at temperatures even higher than *Thermus*. These came to be known as “hyperthermophiles,”



and were shown by their ribosomal RNA sequences to belong to a third kingdom, or domain, of life of Earth: the Archaea. These proved to be sources of polymerases that were even more thermostable.

For example, *Pyrococcus furiosus* (*Pfu*), a hyperthermophilic archaeon, was discovered in the Lower Geyser Basin of Yellowstone National Park (Brock and Freeze, 1969; Brock and Edwards, 1970). Its DNA polymerase (*Pfu*) has been used in many PCR applications. Crystallographic analysis of the native form (Kim et al., 2008) as well as a variant able to replace dCTP with a cyanine dye-labeled dCTP (Wynne et al., 2013) showed that it contains five distinct domains, called the finger, palm, thumb, N-terminal and exonuclease domains (Hopfner et al., 1999; Hashimoto et al., 2001).

Pyrococcus furiosus has a feature that is absent in *Taq* polymerase: an exonuclease domain that has 3'-5' exonuclease activity. This allows *Pfu* to proofread using a conformational change (Hopfner et al., 1999). When the polymerase encounters a mismatch, it binds more weakly to the primer/template, causing strand unwinding. This allows the mismatch to move into the exonuclease pocket, where excision ensues (Freemont et al., 1988). A conserved loop in the exonuclease domain interacts with the thumb domain (Kuroita et al., 2005). Mutation of a key residue H147 to a glutamate residue in this loop results in an electrostatic attraction of the thumb domain to the exonuclease domain, preventing the 3' end of single stranded DNA from entering the exonuclease domain, thus significantly reducing the 3'-5' exonuclease activity (Wang et al., 1997; Kuroita et al., 2005). Kim et al. (2008) suggested that an alternative residue E148 was located at a better position in the loop to interact with the thumb

domain through a comparison of the crystal structures of *Pfu* and KOD1, a homologous Family B polymerase from the *Thermococcus* genus.

Family B DNA polymerases of hyperthermophilic archaeons may generally have 3'-5' exonuclease activity (Joyce, 1989; Joyce and Steitz, 1994; Benkovic et al., 2001; Joyce and Benkovic, 2004). Some have shown to recognize the presence of uracil and hypoxanthine in a template strand, stalling when they sense it ahead of the extension site (Greagg et al., 1999; Fogg et al., 2002; Connolly, 2009). This may reflect functional adaptation. Both uracil and hypoxanthine are “mistakes” in a DNA sequence, arising via the deamination of cytosine and adenine, respectively. Such deaminations presumably occur more rapidly at the higher temperatures where hyperthermophiles live. If the polymerase, nevertheless, extends further through the incorporation of dNTPs placing the uracil in the +2 position, the resulting outcome is the activation of the proofreading excision of the deaminated base (Connolly, 2009).

The need for additional proofreading in the natural environment may not be so pressing to a biotechnologist. Accordingly, many have altered the proofreading ability of *P. furiosus* by either removing the exonuclease activity for use in error-prone PCR (ePCR; Biles and Connolly, 2004) or increasing the efficiency of ligation-mediated PCR protocols (Angers et al., 2001). Sanger sequencing also requires the elimination of the exonuclease activity, otherwise incorporated ddNTPs would be removed and the sequencing signals would disappear.

PROTEIN ENGINEERING AND DIRECTED EVOLUTION THE “NEXT GENERATION” GOALS

For classical Sanger sequencing, a polymerase need only accept a tagged triphosphate with a 3'-blocking group with modest fidelity. The termination:extension ratio will be adjusted in any case by adjusting the concentrations of the terminating and non-terminating triphosphates, meaning that relative inefficiency of incorporation of the unnatural species is not problematic.

However, as the synthetic biology research paradigm has developed, the demands placed on polymerase performance have increased. Here, polymerases are often called upon to copy DNA and PCR amplify molecules containing unnatural nucleotides, often at multiple sites. Here, the fidelity and (preferably) processivity required by a DNA polymerase to support PCR with unnatural nucleotides must be very high. In addition, the structural differences between a DNA polymerase that makes one error per thousand nucleotides and one error per million can be quite subtle and can arise through geometric differences that would not be necessarily distinguished even in a high resolution crystal structure.

RATIONAL DESIGN

Information from structural biology

Molecular biologists would like to believe that they have command of structural theory to “rationally” design polymerases with new, anticipatable properties. In some cases, this is possible, especially when it involves domain shuffling. This has been productive in improving one feature of polymerases important for a wide range of applications: processivity.

DNA binding factors are known to enhance processivity of many polymerases charged with copying complete microbial genomes. In principle, these might be added to improve the processivity of Pol I polymerases, which (as noted above) do not perform this role naturally. Their addition might also, in principle, be used to enhance the performance of any polymerase or polymerase variant. This addition, however, is not often used in biotechnology because of the complexity of the assembled combination. Indeed, *Taq* polymerase and other enzymes are used without accessory proteins for PCR because of their simplicity, which comes from their physiological roles in lagging strand replication and DNA repair.

The complexity of a multicomponent system would be avoided by directly fusing a processivity domain to the active polymerase domain. Adopting this rationale, Wang et al. (2004) covalently fused the double stranded DNA binding protein Sso7d from *Sulfolobus solfataricus* at the N-terminus of *Taq* polymerase (S-*Taq*) and to the fragment of *Taq* polymerase that results from the deletion of the first 289 amino acids which lacks the exonuclease domain [S-*Taq*(Δ289)]. The average length of primer extension prior to template-primer dissociation with *Taq* (Δ289) was increased from 2.9 to 51 nucleotides in S-*Taq*(Δ289). The full-length *Taq* polymerase, which is intrinsically more processive than *Taq* (Δ289), improves its average primer extension from 22 (*Taq*) to 104 (S-*Taq*) nucleotides (Wang et al., 2004).

In parallel work, Wang et al. (2004) also fused the Sso7d domain to the C-terminus of the polymerase from *P. furiosus*, to give *Pfu* polymerase (*Pfu*-S). As in the case of *Taq* polymerase, the fusion of the Ssod7 domain lead to an increase of the average primer extension, from 6.4 nucleotides for *Pfu* to 55 for *Pfu*-S.

Uses for the more processive (*Pfu*-S) were further realized in 1999 when the crystal structure of *Thermococcus gorganarius* DNA polymerase (*Tgo*) was solved. This structure identified a uracil binding pocket, which is used physiologically to prevent the polymerase from copying a template containing uracil, which arises from deamination of cytosine. This structure directed the construction of mutant forms of *Tgo* and *Pfu* DNA polymerases with reduced uracil stalling (Hopfner et al., 1999; Fogg et al., 2002). To increase the ability to read through uracil in the template, the Ssos7 domain was fused to both (*Pfu*-S) and the high fidelity mutant *Pfu* (V93Q). The result was higher processivity and improved uracil-excision cloning (Nour-Eldin et al., 2006).

Structural biology also provided a domain-swapping rational to increase the processivity of *Taq* polymerase. Here, the thioredoxin binding domain (TBD) of the T3 bacteriophage DNA polymerase was inserted into the thumb domain of *Taq* DNA polymerase, deleting amino acids 480–485 (Davidson et al., 2003). The rationale recognized that the processivity of T7 DNA polymerase increases from 15 to 2000 nucleotides when it forms a complex with *E. coli* thioredoxin. The affinity to the primer-template is also increased 80-fold upon binding to thioredoxin. The polymerase arising from this domain fusion remains thermostable, and has a 20–50 times higher processivity than the original *Taq* polymerase.

Exploiting information from multiple sequence alignments (MSAs) in rational engineering

Polymerases are, of course, widely distributed in the biosphere in homologous form. During their divergent evolution, natural selection superimposed upon random variation has carried out several billion years of “protein engineering” experiments, of a sort. With the explosion of microbial sequencing in the last two decades, the results of these “experiments” can be obtained from a public sequence database. To the extent that these results are not corrupted by sequence error, they provide “evolutionary guidance” to assist laboratory protein engineering (Weinhold et al., 1987).

Evolutionary guidance has been productively applied to engineer polymerases, with Tabor and Richardson (1995) providing a classic example. Seeking to improve the ability of *Taq* DNA polymerase I to accept 2',3'-dideoxynucleoside triphosphates (ddNTPs) for sequencing applications, Tabor and Richardson (1995) examined the sequences of three DNA polymerases from Family A (Braithwaite and Ito, 1993). “Wet” biochemistry had already told them that one of these, that from bacteriophage T7, incorporated ddNTPs better than the two others, polymerases from *E. coli* and *Thermus aquaticus*.

Tabor and Richardson (1995) then constructed a multiple sequence alignment (MSAs) for the three homologous Family A polymerases. They noticed that T7 polymerase had a tyrosine at a site (numbered 526) that is homologous to positions that held a phenylalanine in the *E. coli* and *Taq* polymerases (numbered 762 and 667 respectively). From this comparison, they hypothesized that this single amino acid difference was responsible for the different levels of discrimination against ddNTPs among the three polymerases.

Based on this hypothesis, Tabor and Richardson (1995) replaced the phenylalanine in the *Taq* polymerase by a tyrosine. The result was a variant *Taq* (F667Y) that retained the thermostability of the *Taq* parent but gained improved ability to accept ddNTPs. Similar improvements were seen when the analogous replacement was made in the polymerase from *E. coli*. The mutant *Taq* (F667Y) became one of the first “designed” polymerases to be used in DNA sequencing (Tabor and Richardson, 1995).

Subsequently, Li et al. (1999) studied the crystal structures of Klentaq1, a derivative of *Taq* DNA polymerase that lacks an exonuclease domain. In separate structures, protein crystals binding ddNTPs were observed to have closed ternary complexes, where a conformational change upon substrate binding was associated with a large shift in the position of the side chain of residue 660 in the O helix. Comparing the open and closed structures with ddGTP, Li et al. (1999) concluded that the selective interaction of arginine 660 with the O⁶ and N⁷ atoms of the G nucleobase might provide structural grounds for better incorporation of ddGTP by *Taq* polymerase. Guided by these observations, Li et al. (1999) then replaced amino acids at residue 660 in Klentaq1 already holding the Tabor-Richardson replacement (F667Y) and studied the resulting variants. Among the variants, the double mutant *Taq* (F667Y; R660D) showed superior performance in DNA sequencing architectures that used ddNTPs.

THE NEED FOR DIRECTED EVOLUTION: SEQUENCE LANDSCAPES

While structure, evolutionary comparison, and mechanistic analysis are all important tools in polymerase engineering, it remains a fact that chemical theory is inadequate to predict the exact outcome of any amino acid replacement on the performance of any protein, including polymerases. A degree of “trial and error” is inherent in protein engineering experiments. This, in turn, requires that we consider the size of the “protein sequence space” that might be explored as we set out to modify a protein to allow it to support a specific technological goal.

Background to this concept was presented by Smith (1970) almost a half century ago. We begin by noting that the behavior of all possible proteins of length n with respect to a measurable behavior can be represented by a space in n dimensions, where each dimension can have one of 20 discrete values, representing the 20 natural amino acids. Each protein sequence is represented by a point in that space. Two points are neighbors in that space if one can be converted into another by a single amino acid substitution. Thus, with 20 amino acids, each point in the sequence space has $19n$ neighbors. The measurable behavior is a real number displayed in the $n+1$ dimension.

Different sequences have different functions, and moving from a sequence having a function to another functional sequence can proceed via intermediates that either have or lack function. This is illustrated in **Figure 2** with a word game used by Smith (1970), where functional protein sequences are analogous of strings of letters that have a meaning in English. In Smith’s (1970) analogy, the sequence of letters in the word “WORD” is converted to the sequence of letters in the word “GENE” by exchanging one letter

at the time, with each step in one path having a meaning (WORE, GORE, and GONE). Paths where all intermediates are meaningful are illustrated by solid lines between points on the surface. Other paths proceed via words lacking meaning, as illustrated by broken lines (for example, WOND, GOND, and GEND).

In this example, linguistic “meaning” is equated to fitness, which provides the $n+1$ dimension to the surface, a “fitness landscape” (Wright, 1932). The landscape is represented as a topographic map with peaks marked with a (+) for optimal sequences. The absence of function is depicted as dips, marked with a (–). Smith (1970) proposed that natural evolution evolves along paths only if all intermediates are functional. Non-functional sequences are removed by “purifying” selection. Thus, the only valid pathways to explore a sequence space proceed via functional sequences, just as the evolution of words can proceed only via meaningful words.

Sequence space within a protein framework is vast, but enumerable. For example, a 100-amino acid protein can be arranged in 20^{100} different ways. Typical polymerases, eight times longer, constitute a space with 20^{800} points. Both numbers are astronomical. No experiment can sample this space effectively.

Several features of the fitness landscape influence the ease with which it is searched: (a) the fitness landscape is “smooth,” meaning that a useful protein sequence can be obtained starting at any point on the landscape via a path that encounters only other functional proteins or, if not, then (b) useful functional proteins can be obtained no matter where one starts the search, as the surface has many of them or, if not, then (c) the library is guided so as to start the search in a region of the functional hypersurface

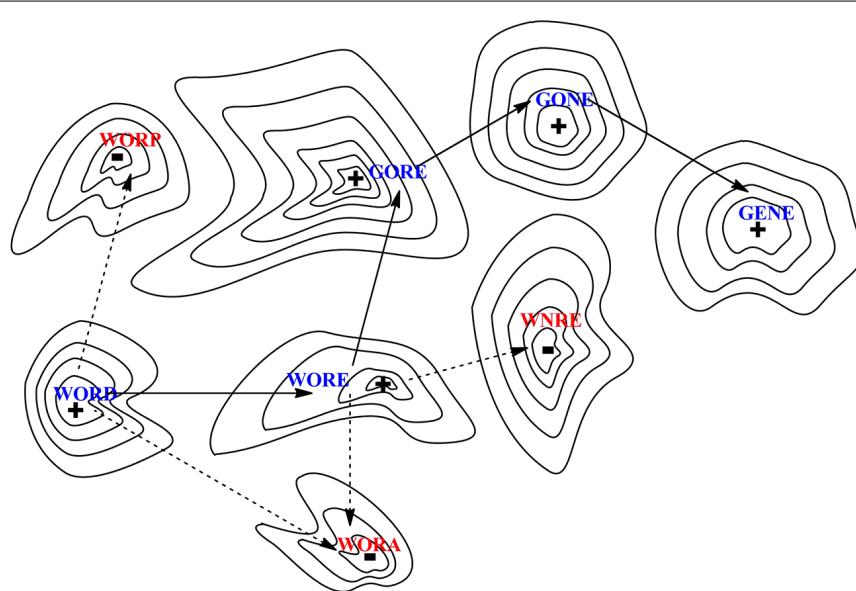


FIGURE 2 | Evolution can be modeled as a walk across a fitness landscape, here presented as a two-dimensional representation of a multiple dimensional hypersurface; analogous to a topographic map, peaks (+) indicate the locations where function exist while dips (–) represent regions with lack of function. Illustrated through an analogy to a word game, a meaningful (functional) string of

letters (here “word”) must be reached starting from another string (“gene”) via stepwise replacement of single letters, where every intermediate along the path must itself also be a functional word. Solid arrows indicate a path of accepted mutations while dashed arrows illustrate deleterious mutations that produce non-functional proteins.

where useful functional proteins reside. Directed evolution is an approach that mimics natural evolution in a time scale that can be reproduced in a laboratory. A directed evolution experiment starts by producing a library of variants (to be discussed further) which then would be selected to a screen or to a selection. A screen involves testing individual variants for the desired properties and is suitable for relatively small libraries, perhaps no more than a few 100s. A selection typically sorts millions of variants at the same time. The experimenter designs the selection in a way that only the variants with the desired properties would survive the selection. The expected outcome of a directed evolution experiment is an enriched pool of variants with proteins having the desired characteristics. Directed evolution can be used to optimize and study any protein (Sterner, 2011).

THE PRACTICE OF DIRECTED EVOLUTION WITH POLYMERASES

In a directed evolution experiment, a “parent” enzyme is chosen to start the search that has (at least) some of the properties desired in the enzyme that will ultimately have utility. The gene of this parent enzyme is then altered to create a library encoding variant forms of the enzyme; some of which might be able to catalyze the desired transformation better than the parent enzyme. The members of the library that are of interest can be isolated by screening or selection.

Library generation

In fact, we have little information about the “smoothness” of any protein fitness landscape. The native polymerase used to initiate an experiment in directed evolution is, of course, already at an elevated point on the fitness landscape, at least for some conditions. It is not clear how many steps (amino acid replacements) can be taken away from the native sequence without losing activity. Further, we expect that certain replacements are more likely to retain core activity than others. All of this suggests that the nature of the library generated from that native sequence might influence the outcome of a directed evolution experiment. It is certainly expected that library generation, if intelligently biased, will allow desired outcomes to be generated faster.

Error prone PCR. A common way to generate libraries from a starting sequence is “mutagenic” or “ePCR.” This approach takes advantage of the inherent propensity of *Taq* polymerase to introduce mistakes into the copies of DNA under certain conditions. The frequency of mismatching is often increased by introducing manganese Mn²⁺ along with the natural cofactor Mg²⁺ (Vartanian et al., 1996). Other additives, such as alcohols or unbalanced concentrations of nucleotides, can also be used to introduce mutations through PCR.

The ePCR method produces does not produce a truly random set of amino acid replacements, for several reasons:

- (i) *Taq* pol tends to replace purines (adenine and guanine) by other purines and pyrimidines (thymidine and cytidine) by other pyrimidines; these changes are called transitions (as opposed to transversions, which exchange a purine for a pyrimidine or a pyrimidine for a purine). The biased tendency of the polymerase to generate transitions over transversions leads to

libraries with amino acid replacements that are non-random with respect to the parent protein.

- (ii) Even if ePCR introduced transitions and transversions equally, the resulting amino acid replacements would not be random, due to the structure of the genetic code. In the code, amino acids having similar chemical properties have closely related codons (Wong et al., 2007). For example, the valine codon (GTN¹) is converted by a single nucleotide replacement to a phenylalanine codon (TTY²), a leucine codon (CTN), an isoleucine codon (ATN), an aspartate codon (GAY) or a glycine codon (GGN). To gain access to codons for other amino acids and, consequently, more dramatically, alter chemical properties in the variant protein, two or three nucleotide replacements are required.

High levels of replacement are not easily achieved by ePCR, nor are they desired. Typical ePCR introduces no more than 4–6 mutations per 1000 nucleotides. Further, a mutation rate that is high enough to search amino acid sequences independent of the code is almost certainly too high to generate any variants that retain polymerase activity as polymerases.

Degenerate codons. Recognizing this challenge, Reetz et al. (2008) developed an elegant approach to library generation that introduces degenerate codons: NNK and NDT. Here, N is any nucleobase, K is guanine or thymine, and D is guanine or adenine or thymine. With the NNK degenerate codon, all 20 amino acids are covered by just 32 ($= 4 \times 4 \times 2$) of the 64 codons possible with standard nucleotides. The twelve NDT degenerate codons ($= 4 \times 3 \times 1$) cover a representative sample of the standard amino acids, including non-polar, aromatic hydrophobic, hydrophilic, and charged amino acids.

Behind this discussion are assumptions about the meaning of the word “random” when discussing amino acid replacements. Some amino acids are encoded by more codons than other amino acids, like serine, with six codons; in contrast, tryptophan is encoded by just one codon. A gene with a truly random sequence would give proteins with a codon-weighted distribution of amino acids. Even this might not be the desired goal of an unguided approach to library generation as some amino acids appear in natural proteins more abundantly than expected from their few codons, for example aspartate and glutamate, each with two codons. Thus, an “ideal” library might arguably be one in which amino acids are replaced by a process that leaves the naturally observed overall composition of the protein unchanged. Finally, our ignorance on the shape of function landscapes, as well as our ignorance of the local topography around any individual parent sequence, means that we cannot state *a priori* which amino acid distribution is most likely to give a desired result in a directed evolution experiment.

Libraries made by gene shuffling or molecular breeding. Random mutagenesis of a parent gene fails, of course, to use all of the information available to a protein engineer, especially in a post-genomic world. As noted above, Nature has already run evolution

¹N is any nucleotide.

²Y is a pyrimidine.

experiments. These provide to us many homologs of a parent protein having many amino replacements relative to the parent sequences. Most of these are functional, and, therefore, identify points in sequence space that are elevated on the fitness landscape. It would be desirable to use the information that these homologs provide.

Gene shuffling was introduced by Stemmer (1994) more than a decade ago to directly use these homologs. Here, the starting point is a family of genes that share enough sequence similarity that they can undergo homologous recombination. Using a modified PCR protocol, gene chimeras are produced.

Those using shuffling in protein evolution assume, of course, that sequence space is more efficiently searched by combining the outcomes of two historically successful searches of a particular region of sequence space, than a search that simply replaces single amino acids starting from a single parent. These historical searches delivered the two functioning proteins whose genes are being shuffled. Here, the landscape is assumed to be such that specific paths between two elevated points are also similarly elevated.

This would be a more compelling hypothesis if natural evolution were observed to use shuffling. Natural evolution does, of course, have access to mechanisms that shuffle parts of genes. Natural evolution uses these mechanisms to rearrange (for example) the order of independently folded units in multi-unit polypeptides. This is famously done in the evolution of multi-unit proteins involved in metazoan signal transduction, where a regulatory protein might contain one “src homology domain 1” unit (SH1, a protein kinase), a few SH2 units, and a few SH3 units (Benner et al., 1993). Evolutionary analysis shows that these are all obtained by shuffling, implying that shuffling is an efficient way to search sequence space when no protein folding unit is disrupted.

However, natural evolution does not provide many examples where polypeptide chains within a *single* folded unit are shuffled. This is presumably because the buried contacts binding collections of secondary structural units are finely tuned to permit packing. Changing a single hydrophobic side chain in a packed protein fold often converts a core that is (typically) as densely packed as an organic crystal into a “molten globule.” Thus, these biophysical realities would make it surprising to expect that shuffling explores sequence space more effectively than point mutation. Such expectations rely, of course, on the view that natural evolution exploits the most effective ways to search sequence space.

Use of evolutionary information to create smaller but better libraries. Alternative approaches now exist to create libraries that search sequence space around parent sequences (Lutz and Patrick, 2004; Jackel et al., 2008; Lutz, 2010). One class of these exploits evolutionary guidance. For example, Cole and Gaucher (2011) introduced an approach, called the Reconstructing Evolutionary Adaptive Paths (REAP) to create libraries that were hypothesized to explore local sequence space with more efficiency. REAP begins with a phylogenetic analysis of homologous sequences, seeking signatures of functional divergence. An amino acid at a site may be entirely conserved in one branch of a phylogenetic tree, while not conserved at all in a second branch. This pattern of divergence, sometimes called *heterotachy*, indicates that the purifying

selective pressures operating in the first branch at this site are different and stronger than those in the second. This, in turn, means that the function of the proteins within the first phylogenetic branch is different from the function in the second branch.

Only rarely, however, has natural history sought a phenotype desired by a protein engineer, of course, only rarely responsive to the specific adaptive changes needed by today’s biotechnologist. Ancient polymerases, for example, were most likely *not* evolving to become resistant to heparin, a target of one of Holliger’s selections. Therefore, the rationale for exploiting “evolutionary guidance” is more subtle.

A REAP analysis identifies sites that have been historically involved in *some* adaptive event. Because *some* changes are involved, the amino acid at the site cannot be *absolutely* required for core function. Conversely, the REAP-identified sites are not likely to be those whose amino acids *never* have a phenotypic impact. The rationale being that sites that have in the past been involved in an adaptive event without losing core function are sites that might be productively examined to identify sites that might adapt the protein to the *new*, biotechnologist-demanded, function.

Thus, the rationale behind REAP is the hypothesis that the most productive sites to replace in a protein engineering experiment are neither sites whose amino acids contribute to a core function (as indicated by their absolute conservation) nor sites in which the choice of amino acid is incidental to function (as indicated by their easy variability). By identifying sites for which replacement might have phenotypic impact without destroying core function, REAP is proposed to have an advantage compared to other methods in the generation of libraries with productively altered behaviors. The advantage of the REAP approach relies on the fact that nature has already tested several amino acid sites, and these modifications on these sites produce enzymes that retain the original activity. Searching for new variants in a REAP library gives the advantage of having several parent enzymes.

Thus, the design of a high fidelity DNA polymerase from a medium fidelity polymerase is largely beyond current structure theory. This makes it impossible to get polymerases with the desired high level behaviors from fully guided protein engineering. As a consequence, many investigators use protein engineering to select for polymerases with certain properties improved with respect to a desired function, starting from libraries of polymerase variants. The directed evolution approach is today considered by many to be the method of choice for protein engineering (Bornscheuer and Pohl, 2001; Yuan et al., 2005; Leemhuis et al., 2009; Turner, 2009).

Compartmentalization

Directed evolution requires the connecting of a phenotype with a genotype in a way that allows only genes that confer a desired phenotype to be propagated. This can be done in many ways. One method is compartmentalized self replication (CSR). Developed by Tawfik and Griffiths (1998), CSR holds proteins and genes together in water droplets suspended in oil emulsions. These generally receive the gene-protein pair from a single *E. coli* cell that is encapsulated

within individual droplets (Tawfik and Griffiths, 1998). When the protein is a polymerase variant, its gene is copied only if that variant is active under the conditions of the evolution experiment.

Compartmentalized self replication was first applied to the directed evolution of DNA polymerases by Ghadessy et al. (2001). Here, a library of polymerase genes was delivered in plasmids to create clones in *E. coli* cells. These cells were dispersed into emulsified water droplets containing the primers and buffers needed to perform a PCR amplification of the polymerase gene. Approximately $\sim 10^8\text{--}10^9$ compartments are formed per milliliter of emulsion; ideally, each compartment contains a single variant. PCR cycling is then performed, with the first heat step lysing the *E. coli* cell to present its expressed thermostable polymerase and its encoding plasmids to the primers. Lysis of the cells then delivers polymerase variants expressed inside of the cells to the buffer, which contains the necessary components for PCR. The polymerase variants and the contents of the buffer remain encapsulated during the PCR cycling.

Polymerases that functioned under the conditions imposed by the experiment were able to make copies of only their own genes. After 20 rounds or more of PCR, the emulsions are broken to give a pool of PCR products enriched in the genes that encoded the selected polymerase variants. These genes could be used directly, or be introduced in cells for another round of

selection. This process is shown schematically in **Figure 3**. With iteration, this process mimics natural evolution, except that the selective pressures applied come from the bioengineer, rather than from Nature.

Phage display is an alternative way to connect genotype and phenotype. In it, a polymerase is linked to its encoding gene in a single viral particle. The protein of interest is co-expressed on the coat of a virus, linking genotype to phenotype. The Romesberg laboratory has been especially active in generating polymerase variants using this approach (Xia et al., 2002; Leconte et al., 2005, 2010).

EXAMPLES OF MODIFIED POLYMERASES

CONVERTING A DNA POLYMERASE TO AN RNA POLYMERASE

Misincorporation by a DNA polymerase through the incorporation of ribonucleoside triphosphates, rather than deoxynucleoside triphosphates, would circumvent the normal pathways in living cells. Accordingly, all DNA polymerases utilize a common mechanism to avoid misincorporation of ribonucleotides by a single active site residue known as the “steric gate” (Joyce, 1997; Gardner and Jack, 1999; Brown and Suo, 2011). Mutations in the steric gate alone are sufficient to render the DNA polymerase able to incorporate nucleoside triphosphates. Yet, products lengths have not exceeded 58 nucleotides and generally result in short termination sequences stalling at +6–7 nucleotides (Gao et al., 1997;

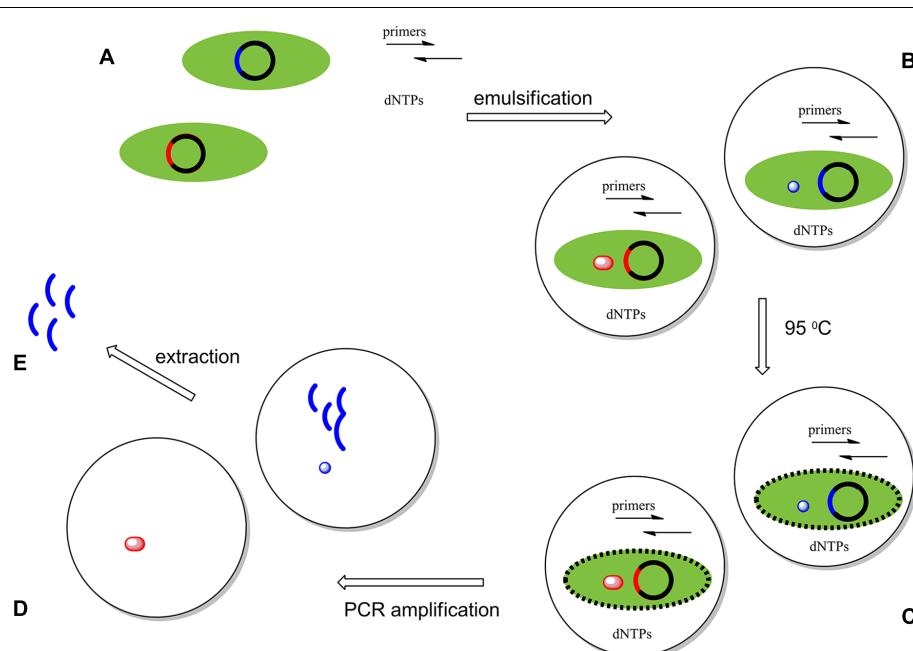


FIGURE 3 | Compartmentalized self replication (CSR) system experiments start with the creation of a library of genes encoding variants of a polymerase. Members of this library are introduced into *E. coli* cells by electroporation. Here, just two variant genes (red and blue) are represented. These genes drive the expression of mutant polymerases in each *E. coli* cell, each of which is isolated in its own water-in-oil-emulsion droplet. **(B)** The first cycle of PCR breaks the cell wall of the *E. coli*, exposing the expressed polymerase molecules and their gene to the contents of a water droplet containing all of the necessary components necessary for a

PCR amplification: (i) primers, (ii) dNTPs, (iii) a mutated gene of the polymerase, and (iv) the enzyme expressed by this gene (**C**). During PCR, any polymerases active under the selective pressure (blue) amplify their respective genes, enriching the pool of mutants having the desired properties; inactive polymerases (red) fail to do so (**D**). The emulsion is then broken and the amplified genes enriched in those encoding polymerases having the desired behaviors are extracted and inserted in a plasmid vector [circular DNA; **E**]. These then enter the cycle of selection again (**A**). After repeating these cycles an enriched pool of variants of the original gene are produced.

Gardner and Jack, 1999; Patel and Loeb, 2000b; Xia et al., 2002; Yang et al., 2002; Ong et al., 2006; McCullum and Chaput, 2009; Brown et al., 2010; Staiger and Marx, 2010; Brown and Suo, 2011). Recently, Cozens et al. (2012) discovered a single amino acid mutation (E664K) in the DNA polymerase from *Thermococcus gorgonarius* that in conjunction with a “steric gate mutation” produced a DNA polymerase capable of synthesizing long RNAs, up to 1.7 kb.

Using phage display the Romesberg laboratory has evolved a DNA polymerase [the Stoffel fragment (Sf) of Taq polymerase] into a RNA polymerase. With just five mutations, one of them the “steric gate mutation” (E615G in Taq) the DNA polymerase was able to incorporate ribonucleotides triphosphates (rNTPs) with rates increased by 10^3 – 10^4 fold compared to the wild type polymerase (Xia et al., 2002). The Holliger laboratory known for the use of the CSR approach has produced a variant of *Taq* polymerase that can incorporate both dNTPs and rNTPs; this variant has only four mutations one of them the “steric gate mutation” mentioned in the previous example (Ong et al., 2006).

DNA POLYMERASES ABLE TO BYPASS DEFECTS

d’Abbadie et al. (2007) shuffled the genes of the polymerases from three *Thermus* species (*aquaticus*, *thermophilus*, and *flavus*) to generate libraries to start a directed evolution experiment to identify DNA polymerases that can extend single, double and quadruple mismatches, process non-canonical primer-template duplexes, and bypass hydantoins and abasic sites (d’Abbadie et al., 2007). They applied these to PCR-amplify cave bear DNA from remains ca. 50 000 years old. These experiments showed that the polymerases obtained by directed evolution applied to these libraries outperformed Taq DNA polymerase and were, therefore, better able to solve a biotechnological problem, here, the sequencing of ancient damaged genomes.

DNA POLYMERASES ABLE TO ACCEPT EXPANDED GENETIC ALPHABETS

One of the AEGIS base pair created in our laboratories is formed between the nucleotides trivially called Z and P (Figure 4). The Z:P pair has a standard Watson-Crick geometry joined by three hydrogen bonds, differing from the standard C:G pair in the arrangement of donor and acceptor groups that form the connecting hydrogen bonds. Both nucleobases place electron density into the minor groove, a density that can accept a hydrogen bond from a polymerase (Geyer et al., 2003). These features allow polymerases to accept dZTP and dPTP as substrates to form duplexes containing Z:P pairs in primer extension reactions, PCR and nested PCR architectures.

In order to improve polymerases able to accept these AEGIS components, we did a selection among $\Delta(1\text{--}279)$ *Taq* using CSR. Two of the best variants identified: variant (M444V/P527A/D551E/E832V) and variant (N580S/L628V/E832V) showed to pause less when challenged *in vitro* to incorporate dZTP opposite P in a template (Laos et al., 2013). Interestingly, our library was created by introducing random mutations on the *Taq* gene, but the outcome of the selection produced variants which contain several sites that have displayed heterotachy

(different rates of change) in their natural history (Lopez et al., 2002). Heterotachy is a sequence pattern such that the rate of evolution acting at an individual site can be slow in one portion of the phylogeny while the rate at the same site can be rapid in a different portion of the phylogeny. Such patterns arise from shifts in the selective constraints acting at individual sites throughout the evolutionary history of a gene family, and by extension, the precise biomolecular behaviors of the homologous proteins are not identical across the phylogeny (Chen et al., 2010; Cole and Gaucher, 2011) suggesting that these sites were involved in an adaptive change in natural polymerase evolution.

The Romesberg laboratory, using phage display, has produced polymerases having an improved ability to incorporate the self-pairing hydrophobic nucleobases analog propynylisocarbostyril (PICS; Leconte et al., 2005).

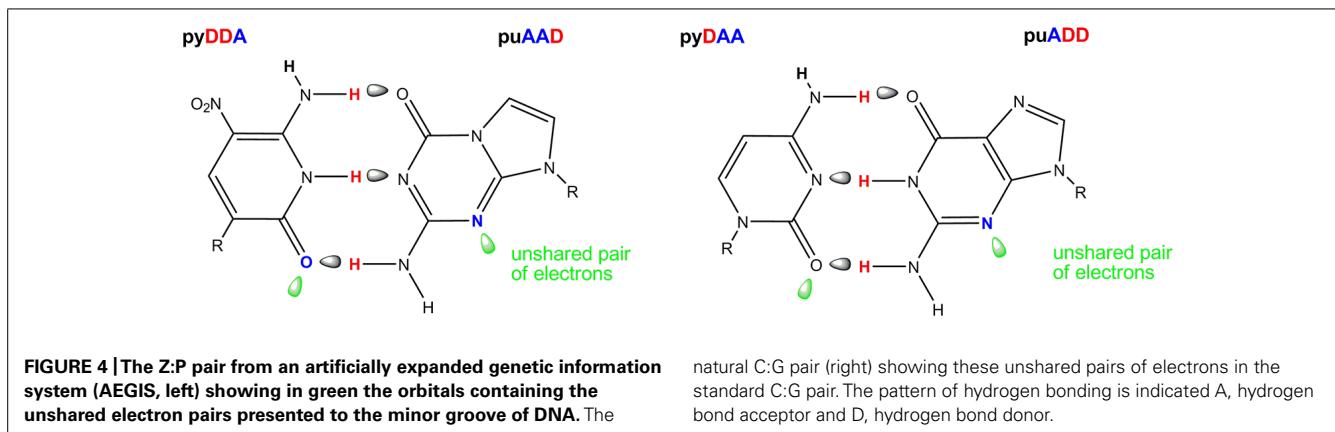
Loakes et al. (2009) have produced a novel polymerase product of the shuffling of polymerases from family A (*Taq* from *T. aquaticus*, *Tth* from *T. thermophilus*, and *Tfl* from *T. flavus*) all of them from the genus *Thermus* and selected by CSR.

POLYMERASES FOR SEQUENCING BY SYNTHESIS METHODS

Sequencing by synthesis (SBS) is a promising next-generation DNA sequencing approach. There are currently several commercial instruments that are offered in the market. Some of the common features of these products are the use of solid phase chemistry to amplify the initial sample and the use of reversible terminators. Reversible terminators are nucleotides that generally have two modifications: one at the 3’OH and the other is either at the 5 or 7 position of the nucleobase. The 3’ hydroxyl position has a cleavable moiety that terminates the polymerase extension reaction after a single-base incorporation. Yet, some reversible terminators do not have a modification on the 3’ hydroxyl like some scarless photocleavable terminator of LaserGen (Wu et al., 2007), the virtual terminator of Helicos BioSciences (Bowers et al., 2009) and the more recently reported Lightning Terminators™ developed in New England Biolabs (Gardner et al., 2012).

The other modification is at the C-5 of pyrimidines or the N-7 position of purines and consist of a fluorescent molecule that is used as a reporter for each of the individual bases. The C-5 and N-7 positions are used because these positions point away from the catalytic pocket of the enzyme. Gardner and Jack (1999) studied variants of Vent DNA polymerase from the hyperthermophilic archaeon *Thermococcus litoralis*. They studied variants on a Tyrosine residue that is highly conserved on family B and was proposed to act as a steric gate (Gardner and Jack, 1999).

The Romesberg laboratories have found polymerases having an improved ability to incorporate modified dUTP with a fluorophore (dUTP-Fl) that can be used for SBS. Leconte et al. (2010) generated a library of Sf, which is *Taq* DNA polymerase minus the first 289 amino acids. This fragment conserves the polymerase activity but lacks the exonuclease domain. The library was done by shuffling the genes of six homologous polymerases: *Thermus aquaticus*; *Thermus thermophilus*; *Thermus caldophilus*; *Thermus filiformis*; *Spirochaeta thermophila*; and *Thermomicrobium roseum*. The three most active polymerase mutants were: Sf168 (with 19 mutations); Sf197 (with 14 mutations). These mutants



had between 10 to 50-fold increase in efficiency for dUTP-Fl incorporation compared with wild type Sf (Leconte et al., 2010).

Our laboratories have produced a variant of *Taq* polymerase using an evolutionary approach to design a polymerase library and then screen a relatively small library for polymerases able to accept unnatural triphosphates modified on their sugar units. Using REAP, they identified 35 sites having heterotachous behavior, after filtering for sites where additional information from evolutionary history, structural biology, and experiments was exploited. They then asked which replacements improve the ability of *Taq* polymerase to accept reversible terminating triphosphates, where the 3'-OH unit of the nucleotide triphosphate had been replaced by an -ONH₂ unit, which prevents continued primer extensions. A single modification (L616A) appears to open space behind Phe-667, allowing the enzyme to accommodate a larger 3'-substituent (Chen et al., 2010).

The Holliger lab, when selecting for variants that accepted 2'-deoxycytidine derivatives carrying appended Cy3- and Cy5 fluorescent dyes, recovered variants of *Pfu* DNA polymerase each having two to six amino acid replacements (Ramsay et al., 2010).

COMMERCIAL APPLICATIONS

The number of commercial applications for non-standard nucleotides is large and growing, implying a growing need for engineered polymerases. This review cannot describe all of the potential commercial applications, but a brief summary of those that already exist indicates their scope. For example, Sherrill et al. (2004) used isoC and isoG modified with reporter molecules to develop an assay to detect both RNA and DNA. These modified nucleic acids (isoC and isoG) are also used to quantify levels of HIV and hepatitis viruses in patients (Collins et al., 1997; Elbeik et al., 2004a,b). isoC and isoG are also used to diagnose a panel of respiratory diseases (Nolte et al., 2007).

Sequencing by synthesis technology relies on nucleoside derivatives that are modified in two ways, first with a fluorescent tag, and then (usually) with a reversibly terminating blocking group (Fedorco et al., 2006; Turcatti et al., 2008). Different modified polymerases, which are commercially available have been suggested to improve the procedure (Aird et al., 2011; Fisher et al.,

2011; Quail et al., 2012). Real-time sequencing also requires polymerases that accept nucleoside derivatives (Eid et al., 2009).

At least one polymerase obtained by directed evolution is commercially available; it was selected to the ability to incorporate dZTP opposite dP in a template, and is available through Firebird Biomolecular Sciences LLC (www.firebirdbio.com).

CONCLUSION AND PERSPECTIVES

The demand for polymerases capable of incorporating unnatural nucleotides is certain to grow as the interest to build modified DNA structures continues, including alternative genetic alphabets (Geyer et al., 2003), highly tagged substrates (Hollenstein et al., 2009), modified backbones (Pinheiro et al., 2012), and other unusual structures (Fa et al., 2004; Leconte et al., 2005; Hirao et al., 2007).

The literature teaches that in some cases, simple downstream screening can obtain polymerases with the needed properties. This is illustrated by efforts by Tabor and Richardson (1995), Gardner and Jack (1999), and Chen et al. (2010) to name a few examples. In these cases to create polymerases that accept various 3'-terminating groups. Their combination of structural biology and evolutionary biology analyses were sufficiently powerful to ensure that regions of sequence space small enough to be screened containing polymerases having the desired properties. In each case, screening began with a relatively small number of variants extracted from the sequence space local around a deftly chosen parent, allowing to get useful enzymes by inspection. However, as a result of the large space sequence space of proteins, examples of success in designing polymerases are not frequent, and are rarely (if ever) *de novo*.

In our experience, the outcome of directed evolution experiments can be explained by the analysis of the evolutionary history of the protein; in this case we found the heterotachy pattern. The heterotachy analysis was originally used to elaborate a small library and screen for variants of *Taq* polymerase (Chen et al., 2010). Later we found several of the substitutions recovered in our directed evolution experiment with *Taq* polymerase had this pattern.

It is interesting to note that some mutations reported in the literature by other research groups happen in amino acid sites considered to be heterotachous by our analysis. One particular amino acid change found by our selection of polymerases

better able to synthesize duplexes containing Z:P pairs (D578N) occurred in a site that underwent substitution in the CSR experiment that obtained a *Taq* variant resistant to heparin inhibition (D578G; Ghadessy et al., 2001). Remarkably, position 614 on *Taq* polymerase has been reported at least three times as the outcome of directed evolution experiments (Patel et al., 2001; Xia et al., 2002; Fa et al., 2004). Other amino acid sites from *Taq* polymerase considered heterotachous and reported on directed evolution experiments are: D144, F598 (Ghadessy et al., 2004), A597, A600, E615 (Xia et al., 2002), and L616 (Patel et al., 2001). This provides support for the general hypothesis behind REAP, that sites involved in adaptation to one environmental novelty might also help adaptation to environmental novelties more generally.

We believe that the recapitulation of the natural history of proteins reflects the fact that the sites can be changed to meet new challenges presented to polymerases without damaging the catalytic power or fidelity of the proteins. The observations in the literature underline the importance of understanding the evolution of polymerases in designing libraries to better explore their sequence space. It will be interesting to further study the outcome of contemporary *in vitro* selection experiments and how they recapitulate.

For less effectively constructed libraries of variants, including those generated by shuffling and by undirected mutagenesis, various selection tools stand to pick up where screening cannot possibly go. Here, CSR and phage display have been especially useful. These have yielded polymerases that support the copying of entirely different genetic systems (Pinheiro et al., 2012).

Although some of the mutations found to be useful for altering polymerases to accept unnatural nucleotides fall on or near the conserved motifs and could potentially be rationalized, there are still a number of mutations that cannot be easily explained and their effect could be subtle.

Other approaches have been found to be useful for the evolution of other enzyme systems. For example, neutral drift libraries (Amitai et al., 2007; Bloom et al., 2007a,b), have yet to be applied as starting points for directed evolution experiments with DNA polymerases.

Further, we (and many others) are seeking to develop living systems that implement a “synthetic biology” based on unnatural DNA analogs. These have the potential for being “biosafe” platforms for artificial metabolisms, fermentations, diagnostics, and therapeutic tools, *inter alia* (Schmidt, 2010).

ACKNOWLEDGMENTS

The authors are thankful to Dr. Dietlind Gerloff from FFAME for useful suggestions in the preparation of this manuscript. We are indebted to DTRA for funding through grant HDTRA1-13-1-0004.

REFERENCES

- Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, R18. doi: 10.1186/gb-2011-12-2-r18
- Amitai, G., Gupta, R. D., and Tawfik, D. S. (2007). Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* 1, 67–78. doi: 10.2976/1.2739115
- Angers, M., Cloutier, J. F., Castonguay, A., and Drouin, R. (2001). Optimal conditions to use Pfu exo(-) DNA polymerase for highly efficient ligation-mediated polymerase chain reaction protocols. *Nucleic Acids Res.* 29, e83. doi: 10.1093/nar/29.16.e83
- Benkovic, S. J., Valentine, A. M., and Salinas, F. (2001). Replisome-mediated DNA replication. *Annu. Rev. Biochem.* 70, 181–208. doi: 10.1146/annurev.biochem.70.1.181
- Benner, S. A. (2004). Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res.* 37, 784–797. doi: 10.1021/ar040004z
- Benner, S. A., Cohen, M. A., and Gerloff, D. (1993). Predicted secondary structure for the src homology-3 domain. *J. Mol. Biol.* 229, 295–305. doi: 10.1006/jmbi.1993.1035
- Berdis, A. J. (2014). “DNA polymerases that perform template-independent DNA synthesis,” in *Nucleic Acid Polymerases*, eds K. S. Murakami and M. A. Trakselis (Berlin: Springer), 109–137.
- Biles, B. D., and Connolly, B. A. (2004). Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res.* 32, e176. doi: 10.1093/nar/gnh174
- Bloom, J. D., Lu, Z., Chen, D., Raval, A., Venturelli, O. S., and Arnold, F. H. (2007a). Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol.* 5:29. doi: 10.1186/1741-7007-5-29
- Bloom, J. D., Romero, P. A., Lu, Z., and Arnold, F. H. (2007b). Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* 2, 17. doi: 10.1186/1745-6150-2-17
- Bornscheuer, U. T., and Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* 5, 137–143. doi: 10.1016/s1367-5931(00)00182-184
- Bowers, J., Mitchell, J., Beer, E., Buzby, P. R., Causey, M., Efcavitch, J. W., et al. (2009). Virtual terminator nucleotides for next-generation DNA sequencing. *Nat. Methods* 6, 593–595. doi: 10.1038/nmeth.1354
- Braithwaite, D. K., and Ito, J. (1993). Compilation, alignment, and phylogenetic-relationships of dna-polymerases. *Nucleic Acids Res.* 21, 787–802. doi: 10.1093/nar/21.4.787
- Brock, T. D., and Edwards, M. R. (1970). Fine structure of *Thermus aquaticus*, An Extreme thermophile. *J. Bacteriol.* 104, 509–517. doi: 10.1073/pnas.0712334105
- Brock, T. D., and Freeze, H. (1969). *Thermus aquaticus* gen N and sp N A nonsporulating extreme thermophile. *J. Bacteriol.* 98, 289–297.
- Brown, J. A., Fiala, K. A., Fowler, J. D., Sherrer, S. M., Newmister, S. A., Duym, W. W., et al. (2010). A novel mechanism of sugar selection utilized by a human X-family DNA polymerase. *J. Mol. Biol.* 395, 282–290. doi: 10.1016/j.jmb.2009.11.003
- Brown, J. A., and Suo, Z. (2011). Unlocking the sugar “Steric Gate” of DNA polymerases. *Biochemistry* 50, 1135–1142. doi: 10.1021/bi101915z
- Cann, I. K. O., and Ishino, Y. (1999). Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* 152, 1249–1267.
- Chen, C. Y. (2014). DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front. Microbiol.* 5:305. doi: 10.3389/fmicb.2014.00305
- Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., et al. (2010). Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1948–1953. doi: 10.1073/pnas.0908463107
- Chien, A., Edgar, D. B., and Trela, J. M. (1976). Deoxyribonucleic-acid polymerase from extreme thermophile *Thermus-aquaticus*. *J. Bacteriol.* 127, 1550–1557.
- Cole, M. F., and Gaucher, E. A. (2011). Exploiting models of molecular evolution to efficiently direct protein engineering. *J. Mol. Evol.* 72, 193–203. doi: 10.1007/s00239-010-9415-9412
- Collins, M. L., Irvine, B., Tyner, D., Fine, E., Zayati, C., Chang, C. A., et al. (1997). A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res.* 25, 2979–2984. doi: 10.1093/nar/25.15.2979
- Connolly, B. A. (2009). Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochem. Soc. Trans.* 37, 65–68. doi: 10.1042/bst0370065
- Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R., and Holliger, P. (2012). A short adaptive path from DNA to RNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8067–8072. doi: 10.1073/pnas.1120964109
- d’Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., et al. (2007). Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* 25, 939–943. doi: 10.1038/nbt1321

- Davidson, J. F., Fox, R., Harris, D. D., Lyons-Abbott, S., and Loeb, L. A. (2003). Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of Taq DNA polymerase. *Nucleic Acids Res.* 31, 4702–4709. doi: 10.1093/nar/gkg667
- Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* 3, 461–467. doi: 10.1093/protein/3.6.461
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138. doi: 10.1126/science.1162986
- Elbeik, T., Markowitz, N., Nassos, P., Kumar, U., Beringer, S., Haller, B., et al. (2004a). Simultaneous runs of the bayer VERSANT HIV-1 version 3.0 and HCV bDNA version 3.0 quantitative assays on the system 340 platform provide reliable quantitation and improved work flow. *J. Clin. Microbiol.* 42, 3120–3127. doi: 10.1128/jcm.42.7.3120-3127.2004
- Elbeik, T., Surtihadi, J., Destree, M., Gorlin, J., Holodniy, M., Jortani, S. A., et al. (2004b). Multicenter evaluation of the performance characteristics of the Bayer VERSANT HCV RNA 3.0 assay (bDNA). *J. Clin. Microbiol.* 42, 563–569. doi: 10.1128/jcm.42.2.563-569.2004
- Eom, S. H., Wang, J. M., and Steitz, T. A. (1996). Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* 382, 278–281. doi: 10.1038/382278a0
- Fa, M., Radeghieri, A., Henry, A. A., and Romesberg, F. E. (2004). Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* 126, 1748–1754. doi: 10.1021/ja038525p
- Fedorco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 34, e22. doi: 10.1093/nar/gnj023
- Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T. M., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 12, R1. doi: 10.1186/gb-2011-12-1-r1
- Fogg, M. J., Pearl, L. H., and Connolly, B. A. (2002). Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* 9, 922–927. doi: 10.1038/nsb867
- Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988). Cocrystal structure of an editing complex of klenow fragment with dna. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8924–8928. doi: 10.1073/pnas.85.23.8924
- Gao, G. X., Orlova, M., Georgiadis, M. M., Hendrickson, W. A., and Goff, S. P. (1997). Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. U.S.A.* 94, 407–411. doi: 10.1073/pnas.94.2.407
- Gardner, A. F., and Jack, W. E. (1999). Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.* 27, 2545–2553. doi: 10.1093/nar/27.12.2545
- Gardner, A. F., Wang, J., Wu, W., Karouby, J., Li, H., Stupi, B. P., et al. (2012). Rapid incorporation kinetics and improved fidelity of a novel class of 3'-OH unblocked reversible terminators. *Nucleic Acids Res.* 40, 7404–7415. doi: 10.1093/nar/gks330
- Geyer, C. R., Battersby, T. R., and Benner, S. A. (2003). Nucleobase pairing in Watson-Crick-like genetic expanded information systems. *Structure* 11, 1485–1498. doi: 10.1016/j.str.2003.11.008
- Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4552–4557. doi: 10.1073/pnas.071052198
- Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., et al. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* 22, 755–759. doi: 10.1038/nbt974
- Greagg, M. A., Fogg, A. M., Panayotou, G., Evans, S. J., Connolly, B. A., and Pearl, L. H. (1999). A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9045–9050. doi: 10.1073/pnas.96.16.9045
- Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., et al. (2001). Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J. Mol. Biol.* 306, 469–477. doi: 10.1006/jmbi.2000.4403
- Henry, A. A., and Romesberg, F. E. (2003). Beyond A, C, G and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.* 7, 727–733. doi: 10.1016/j.cbpa.2003.10.011
- Hirao, I., Harada, Y., Kimoto, M., Mitsui, T., Fujiwara, T., and Yokoyama, S. (2004). A two-unnatural-base-pair system toward the expansion of the genetic code. *J. Am. Chem. Soc.* 126, 13298–13305. doi: 10.1021/ja047201d
- Hirao, I., Mitsui, T., Kimoto, M., and Yokoyama, S. (2007). An efficient unnatural base pair for PCR amplification. *J. Am. Chem. Soc.* 129, 15549–15555. doi: 10.1021/ja073830m
- Hollenstein, M., Hipolito, C. J., Lam, C. H., and Perrin, D. M. (2009). A DNazyme with three protein-like functional groups: enhancing catalytic efficiency of M2+-independent RNA cleavage. *Chembiochem* 10, 1988–1992. doi: 10.1002/cbic.200900314
- Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., et al. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3600–3605. doi: 10.1073/pnas.96.7.3600
- Ishikawa, M., Hirao, I., and Yokoyama, S. (2000). Synthesis of 3-(2-deoxy-beta-D-ribofuranosyl)pyridin-2-one and 2-amino-6-(N,N-dimethylamino)-9-(2-deoxy-beta-D-ribofuranosyl)purine derivatives for an unnatural base pair. *Tetrahedron Lett.* 41, 3931–3934. doi: 10.1016/s0040-4039(00)00520-527
- Ito, J., and Braithwaite, D. K. (1991). Compilation and alignment of dna-polymerase sequences. *Nucleic Acids Res.* 19, 4045–4057. doi: 10.1093/nar/19.15.4045
- Jackel, C., Kast, P., and Hilvert, D. (2008). Protein design by directed evolution. *Annu. Rev. Biophys.* 153–173. doi: 10.1146/annurev.biophys.37.032807.125832
- Joyce, C. M. (1989). How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of dna-polymerase-i (klenow fragment). *J. Biol. Chem.* 264, 10858–10866.
- Joyce, C. M. (1997). Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc. Natl. Acad. Sci. U.S.A.* 94, 1619–1622. doi: 10.1073/pnas.94.5.1619
- Joyce, C. M., and Benkovic, S. J. (2004). DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* 43, 14317–14324. doi: 10.1021/bi048422z
- Joyce, C. M., and Steitz, T. A. (1994). Function and structure relationships in dna-polymerases. *Annu. Rev. Biochem.* 63, 777–822. doi: 10.1146/annurev.bi.63.070194.004021
- Kim, S. W., Kim, D. U., Kim, J. K., Kang, L. W., and Cho, H. S. (2008). Crystal structure of *Pfu*, the high fidelity DNA polymerase from *Pyrococcus furiosus*. *Int. J. Biol. Macromol.* 42, 356–361. doi: 10.1016/j.jbiomac.2008.01.010
- Kim, Y., Eom, S. H., Wang, J. M., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995). Crystal-structure of *Thermus-aquaticus* DNA-polymerase. *Nature* 376, 612–616. doi: 10.1038/376612a0
- Kool, E. T. (2002). Replacing the nucleohashes in DNA with designer molecules. *Acc. Chem. Res.* 35, 936–943. doi: 10.1021/ar000183u
- Kornberg, A. (1960). Biologic synthesis of deoxyribonucleic acid. *Science* 131, 1503–1508. doi: 10.1126/science.131.3412.1503
- Korolev, S., Nayal, M., Barnes, W. M., Dicera, E., and Waksman, G. (1995). Crystal-structure of the large fragment of *Thermus-aquaticus* DNA-polymerase-I At 2.5-angstrom resolution - structural basis for thermostability. *Proc. Natl. Acad. Sci. U.S.A.* 92, 9264–9268. doi: 10.1073/pnas.92.20.9264
- Kuroita, T., Matsumura, H., Yokota, N., Kitabayashi, M., Hashimoto, H., Inoue, T., et al. (2005). Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases. *J. Mol. Biol.* 351, 291–298. doi: 10.1016/j.jmb.2005.06.015
- Laos, R., Shaw, R., Leal, N. A., Gaucher, E., and Benner, S. (2013). Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry* 52, 5288–5294. doi: 10.1021/bi400558c
- Leconte, A. M., Chen, L., and Romesberg, F. E. (2005). Polymerase evolution: efforts toward expansion of the genetic code. *J. Am. Chem. Soc.* 127, 12470–12471. doi: 10.1021/ja053322h
- Leconte, A. M., Patel, M. P., Sass, L. E., Mcinerney, P., Jarosz, M., Kung, L., et al. (2010). Directed evolution of DNA polymerases for next-generation sequencing. *Angew. Chem. Int. Ed. Engl.* 49, 5921–5924. doi: 10.1002/anie.201001607
- Leemhuis, H., Kelly, R. M., and Dijkhuizen, L. (2009). Directed evolution of enzymes: library screening strategies. *IUBMB Life* 61, 222–228. doi: 10.1002/iub.165
- Le Grice, S. F. J., and Nowotny, M. (2014). “Reverse Transcriptases,” in *Nucleic Acid Polymerases*, eds K. S. Murakami and M. A. Trakselis (New York, NY: Springer-Verlag Berlin Heidelberg), 189–214. doi: 10.1007/978-3-642-39796-7_8

- Li, Y., Korolev, S., and Waksman, G. (1998). 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.* 17, 7514–7525. doi: 10.1093/emboj/17.24.7514
- Li, Y., Mitaxov, V., and Waksman, G. (1999). Structure-based design of *Taq* DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9491–9496. doi: 10.1073/pnas.96.17.9491
- Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T., and Holliger, P. (2009). Evolving a polymerase for hydrophobic base analogues. *J. Am. Chem. Soc.* 131, 14827–14837. doi: 10.1021/ja9039696
- Lopez, P., Casane, D., and Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19, 1–7. doi: 10.1093/oxfordjournals.molbev.003973
- Lutz, S. (2010). Beyond directed evolution-semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* 21, 734–743. doi: 10.1016/j.copbio.2010.08.011
- Lutz, S., and Patrick, W. M. (2004). Novel methods for directed evolution of enzymes: quality, not quantity. *Curr. Opin. Biotechnol.* 15, 291–297. doi: 10.1016/j.copbio.2004.05.004
- McCullum, E. O., and Chaput, J. C. (2009). Transcription of an RNA aptamer by a DNA polymerase. *Chem. Commun.* 2938–2940. doi: 10.1039/b820678c
- Minakawa, N., Kojima, N., Hikishima, S., Sasaki, T., Kiyosue, A., Atsumi, N., et al. (2003). New base pairing motifs. The synthesis and thermal stability of oligodeoxynucleotides containing imidazopyridopyrimidine nucleosides with the ability to form four hydrogen bonds. *J. Am. Chem. Soc.* 125, 9970–9982. doi: 10.1021/ja0347686
- Minakawa, N., Kuramoto, K., Hikishima, S., and Matsuda, A. (2006). Four hydrogen-bonding motifs in oligonucleotides. *Arkivoc* 2006, 326–337. doi: 10.3998/ark.5550190.0007.724
- Moon, A. F., Garcia-Diaz, M., Batra, V. K., Beard, W. A., Bebenek, K., Kunkel, T. A., et al. (2007). The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair* 6, 1709–1725. doi: 10.1016/j.dnarep.2007.05.009
- Morales, J. C., and Kool, E. T. (1999). Minor groove interactions between polymerase and DNA: more essential to replication than Watson-Crick hydrogen bonds? *J. Am. Chem. Soc.* 121, 2323–2324. doi: 10.1021/ja983502+
- Nolte, F. S., Marshall, D. J., Rasberry, C., Schielbein, S., Banks, G. G., Storch, G. A., et al. (2007). MultiCode-PLx system for multiplexed detection of seventeen respiratory viruses. *J. Clin. Microbiol.* 45, 2779–2786. doi: 10.1128/jcm.00669-667
- Nour-Eldin, H. H., Hansen, B. G., Norholm, M. H. H., Jensen, J. K., and Halkier, B. A. (2006). Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* 34, e122. doi: 10.1093/nar/gkl635
- Ohmori, H., Friedberg, E. C., Fuchs, R. P. P., Goodman, M. F., Hanaoka, F., Hinkle, D., et al. (2001). The Y-family of DNA polymerases. *Mol. Cell.* 8, 7–8. doi: 10.1016/s1097-2765(01)00278-277
- Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA-polymerase-i complexed with dTMP. *Nature* 313, 762–766. doi: 10.1038/313762a0
- Ong, J. L., Loakes, D., Jaroslawski, S., Too, K., and Holliger, P. (2006). Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* 361, 537–550. doi: 10.1016/j.jmb.2006.06.050
- Patel, P. H., Kawate, H., Adman, E., Ashbach, M., and Loeb, L. K. (2001). A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J. Biol. Chem.* 276, 5044–5051. doi: 10.1074/jbc.M008701200
- Patel, P. H., and Loeb, L. A. (2000a). DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5095–5100. doi: 10.1073/pnas.97.10.5095
- Patel, P. H., and Loeb, L. A. (2000b). Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* 275, 40266–40272. doi: 10.1074/jbc.M005757200
- Piccirilli, J. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990). Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 343, 33–37. doi: 10.1038/343033a0
- Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., et al. (2012). Synthetic genetic polymers capable of heredity and evolution. *Science* 336, 341–344. doi: 10.1126/science.1217622
- Pryor, J. M., Dieckman, L. M., Boehm, E. M., and Washington, M. T. (2014). “Eukaryotic Y-family polymerases: a biochemical and structural perspective,” in *Nucleic Acid Polymerases*, eds K. S. Murakami and M. A. Trakselis (New York, NY: Springer-Verlag Berlin Heidelberg), 85–108.
- Quail, M. A., Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., Mcquillan, J. A., et al. (2012). Optimal enzymes for amplifying sequencing libraries. *Nat. Methods* 9, 10–11. doi: 10.1038/nmeth.1814
- Ramsay, N., Jemth, A. S., Brown, A., Crampton, N., Dear, P., and Holliger, P. (2010). CyDNA: synthesis and replication of highly Cy-Dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc.* 132, 5096–5104. doi: 10.1021/ja909180c
- Rappaport, H. P. (1988). The 6-thioguanine 5-methyl-2-pyrimidinone base pair. *Nucleic Acids Res.* 16, 7253–7267. doi: 10.1093/nar/16.15.7253
- Reetz, M. T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the numbers problem in directed evolution. *Chembiochem* 9, 1797–1804. doi: 10.1002/cbic.200800298
- Schmidt, M. (2010). Xenobiology: a new form of life as the ultimate biosafety tool. *Bioessays* 32, 322–331. doi: 10.1002/bies.200900147
- Sefah, K., Yang, Z. Y., Bradley, K. M., Hoshika, S., Jimenez, E., Zhang, L. Q., et al. (2014). In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1449–1454. doi: 10.1073/pnas.1311778111
- Sherrill, C. B., Marshall, D. J., Moser, M. J., Larsen, C. A., Daude-Snow, L., and Prudent, J. R. (2004). Nucleic acid analysis using an expanded genetic alphabet to quench fluorescence. *J. Am. Chem. Soc.* 126, 4550–4556. doi: 10.1021/ja0315558
- Sismour, A. M., and Benner, S. A. (2005). Synthetic biology. *Expert Opin. Biol. Ther.* 5, 1409–1414. doi: 10.1517/14712598.5.11.1409
- Smith, J. M. (1970). Natural selection and concept of a protein space. *Nature* 225, 563–564. doi: 10.1038/225563a0
- Staiger, N., and Marx, A. (2010). A DNA Polymerase with increased reactivity for ribonucleotides and C5-modified deoxyribonucleotides. *Chembiochem* 11, 1963–1966. doi: 10.1002/cbic.201000384
- Stemmer, W. P. C. (1994). Rapid evolution of a protein in-vitro by DNA shuffling. *Nature* 370, 389–391. doi: 10.1038/370389a0
- Stern, R. (2011). Directed evolution: a powerful approach to optimising and understanding enzymes. *Chembiochem* 12, 1439–1440. doi: 10.1002/cbic.201100285
- Suzuki, M., Avicola, A. K., Hood, L., and Loeb, L. A. (1997). Low fidelity mutants in the O-helix of *Thermus aquaticus* DNA polymerase I. *J. Biol. Chem.* 272, 11228–11235. doi: 10.1074/jbc.272.17.11228
- Switzer, C., Moroney, S. E., and Benner, S. A. (1989). Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* 111, 8322–8323. doi: 10.1021/ja00203a067
- Tabor, S., and Richardson, C. C. (1995). A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6339–6343. doi: 10.1073/pnas.92.14.6339
- Tae, E. J. L., Wu, Y. Q., Xia, G., Schultz, P. G., and Romesberg, F. E. (2001). Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. *J. Am. Chem. Soc.* 123, 7439–7440. doi: 10.1021/ja010731e
- Tawfik, D. S., and Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* 16, 652–656. doi: 10.1038/nbt0798-652
- Turcatti, G., Romieu, A., Fedurco, M., and Tairi, A. P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* 36, e25. doi: 10.1093/nar/gkn021
- Turner, N. J. (2009). Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* 5, 568–574. doi: 10.1038/nchembio.203
- Vartanian, J. P., Henry, M., and Wainhobson, S. (1996). Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.* 24, 2627–2631. doi: 10.1093/nar/24.14.2627
- Wang, J., Sattar, A., Wang, C. C., Karim, J. D., Konigsberg, W. H., and Steitz, T. A. (1997). Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell* 89, 1087–1099. doi: 10.1016/s0092-8674(00)80296-2

- Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* 32, 1197–1207. doi: 10.1093/nar/gkh271
- Weinhold, E. G., Ellington, A., Presnell, S. R., Mcgeehan, G. M., and Benner, S. A. (1987). Evolution guidance - engineering alcohol-dehydrogenase and ribonuclease. *Protein Eng.* 1, 236–237.
- Wong, T. S., Roccatano, D., and Schwaneberg, U. (2007). Challenges of the genetic code for exploring sequence space in directed protein evolution. *Biocatal. Biotransformation* 25, 229–241. doi: 10.1080/1024242070144280
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth. Internat. Congr. Genet. Ithaca N. Y.* 1, 356–366.
- Wu, W., Stupi, B. P., Litosh, V. A., Mansouri, D., Farley, D., Morris, S., et al. (2007). Termination of DNA synthesis by N6-alkylated, not 3'-O-alkylated, photocleavable 2'-deoxyadenosine triphosphates. *Nucleic Acids Res.* 35, 6339–6349. doi: 10.1093/nar/gkm689
- Wynne, S. A., Pinheiro, V. B., Holliger, P., and Leslie, A. G. W. (2013). Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly Cy-Dye labelled DNA. *PLoS ONE* 8:e70892. doi: 10.1371/journal.pone.0070892
- Xia, G., Chen, L. J., Sera, T., Fa, M., Schultz, P. G., and Romesberg, F. E. (2002). Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6597–6602. doi: 10.1073/pnas.102577799
- Yamtich, J., and Sweasy, J. B. (2010). DNA polymerase family X: function, structure, and cellular roles. *Biochim. Biophys. Acta.* 1804, 1136–1150. doi: 10.1016/j.bbapap.2009.07.008
- Yang, G. W., Franklin, M., Li, J., Lin, T. C., and Konigsberg, W. (2002). A conserved Tyr residue is required for sugar selectivity in a pol alpha DNA polymerase. *Biochemistry* 41, 10256–10261. doi: 10.1021/bi0202171
- Yuan, L., Kurek, I., English, J., and Keenan, R. (2005). Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* 69, 373–392. doi: 10.1128/jmbr.69.3.393-392.2005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 June 2014; accepted: 07 October 2014; published online: 31 October 2014.

Citation: Laos R, Thomson JM and Benner SA (2014) DNA Polymerases engineered by directed evolution to incorporate non-standard nucleotides. *Front. Microbiol.* 5:565. doi: 10.3389/fmicb.2014.00565

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Laos, Thomson and Benner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP)

Yogesh Chander¹, Jim Koelbl¹, Jamie Puckett¹, Michael J. Moser¹, Audrey J. Klingele¹, Mark R. Liles², Abel Carrias², David A. Mead¹ and Thomas W. Schoenfeld^{1*}

¹ Lucigen Corporation, Middleton, WI, USA

² Department of Biological Sciences, Auburn University, Auburn, AL, USA

Edited by:

Andrew F. Gardner, New England Biolabs, USA

Reviewed by:

Marla Tuffin, University of the Western Cape, South Africa

Daniel M. Jenkins, University of Hawaii, USA

***Correspondence:**

Thomas W. Schoenfeld, Lucigen Corporation, 2905 Parmenter Street, Middleton, WI 53719, USA
e-mail: tschoenfeld@lucigen.com

Meeting the goal of providing point of care (POC) tests for molecular detection of pathogens in low resource settings places stringent demands on all aspects of the technology. OmniAmp DNA polymerase (Pol) is a thermostable viral enzyme that enables true POC use in clinics or in the field by overcoming important barriers to isothermal amplification. In this paper, we describe the multiple advantages of OmniAmp Pol as an isothermal amplification enzyme and provide examples of its use in loop-mediated isothermal amplification (LAMP) for pathogen detection. The inherent reverse transcriptase activity of OmniAmp Pol allows single enzyme detection of RNA targets in RT-LAMP. Common methods of nucleic acid amplification are highly susceptible to sample contaminants, necessitating elaborate nucleic acid purification protocols that are incompatible with POC or field use. OmniAmp Pol was found to be less inhibited by whole blood components typical in certain crude sample preparations. Moreover, the thermostability of the enzyme compared to alternative DNA polymerases (*Bst*) and reverse transcriptases allows pretreatment of complete reaction mixes immediately prior to amplification, which facilitates amplification of highly structured genome regions. Compared to *Bst*, OmniAmp Pol has a faster time to result, particularly with more dilute templates. Molecular diagnostics in field settings can be challenging due to the lack of refrigeration. The stability of OmniAmp Pol is compatible with a dry format that enables long term storage at ambient temperatures. A final requirement for field operability is compatibility with either commonly available instruments or, in other cases, a simple, inexpensive, portable detection mode requiring minimal training or power. Detection of amplification products is shown using lateral flow strips and analysis on a real-time PCR instrument. Results of this study show that OmniAmp Pol is ideally suited for low resource molecular detection of pathogens.

Keywords: diagnostics, RNA/DNA polymerase, infectious diseases, RT-LAMP, point-of-care

INTRODUCTION

Rapid, sensitive, easy-to-use methods for detection of pathogens are needed for timely diagnosis of infectious diseases especially at point-of-care (POC). Common molecular detection methods by end point and real time PCR are valuable tools for pathogen detection and are widely used in clinical diagnostics because of high sensitivity and specificity (Segawa et al., 2014). However, there are several problems in implementing these methods at POC, particularly the need for trained personnel, extensive sample preparation protocols and specialized laboratory equipment, which have prevented use of these methods in resource limited settings. Isothermal amplification methods such as loop mediated amplification (LAMP; Notomi et al., 2000) hold great promise to shorten nucleic acid detection times, simplify the instrumentation and reduce power requirements by eliminating the need for thermal cycling. These improvements are facilitating the movement of nucleic acid tests (NATs) from the central laboratory to

POC environments like clinics, hospital emergency rooms, farms and other remote areas of need, but still require improvement before fulfilling their potential.

While LAMP has proven highly useful in laboratory environments, the current formats have limited application under POC conditions (Nijru, 2012). Improvements in the DNA polymerase and detection modes could allow use of the LAMP platform in POC testing in resource limited settings. A key drawback of typical LAMP formulations is the inability to directly detect RNA without a second reverse transcriptase enzyme. Currently, most LAMP methods use a truncated product (large fragment) of *Bacillus stearothermophilus* (*Bst*) Pol (Huang et al., 1999) or a highly similar enzyme from closely related moderately thermophilic bacterium. While this enzyme is highly effective for amplification of DNA based targets, it cannot amplify RNA without the addition of a reverse transcriptase for conversion of RNA template to cDNA that serves as a target for LAMP. This

adds additional steps and necessitates use of a buffer that is a compromise between the optimal conditions for the respective enzymes. The reaction requirement for the reverse transcriptase also imposes a limit on the thermal stability of the reaction.

Some of the most important RNA targets for diagnostic detection are viral genomes, which can be highly structured. Thermal treatment during sample preparation immediately prior to amplification has been indispensable in allowing direct detection of bacterial and viral targets. Currently available reverse transcriptase's, Avian Myeloblastosis Virus (AMV RT) or Moloney Murine Leukemia Virus (MMLV RT), are relatively labile and thermal melting to alleviate secondary structure has not been possible with any of the current LAMP systems. While single enzyme RT-LAMP methods are available (<http://www.optigene.co.uk/>), most RT LAMP uses the two-enzyme format.

In order to use NATs in POC settings, it is important to have a simple and easy to use method for detecting amplification. Ideally the detection would confer additional specificity and sensitivity, while keeping total testing costs low. Most common detection methods for LAMP, such as agarose gel electrophoresis or use of real-time PCR instruments are prohibitively expensive, slow, and require extensive user training. Dyes such as calcine (Tomita et al., 2008), or hydroxynaphthol blue (HNB; Goto et al., 2009) in the LAMP reaction mixture allows direct visual detection of amplification results, but do not improve specificity and the ambiguous results require more user judgment than is acceptable for POC use. An alternative detection mode is the use of lateral flow devices (LFD), which can be portable, and does not require instrumentation or electrical power. The combination of LAMP and LFD provides an inexpensive, facile tool for NAT in remote, low resource environments. The need for a refrigerated cold chain is unavailable in many low resource settings, which impairs the utility of a POC test, so a final component of a molecular based POC technology is the stability of the test for distribution and storage under ambient conditions.

Screening viral metagenomes from boiling hot springs uncovered new thermostable DNA polymerases (Schoenfeld et al., 2008). An engineered derivative of one of these, PyroPhage 3173

DNA polymerase, was effective in RT PCR (Moser et al., 2012). This enzyme exhibits innate reverse transcriptase activity, thermostability and potent strand-displacing activity and has now been formulated for use in direct detection of RNA and DNA pathogens by LAMP. Its thermostability allows additional flexibility for using a thermal treatment in sample preparation and amplification of highly structured regions of genomes.

In this report we describe the use of this novel polymerase in LAMP and RT-LAMP (reverse transcription LAMP). In order to understand the potential applications and limitations of using OmniAmp polymerase in LAMP, a diverse group of DNA and RNA based targets were selected (Table 1). In addition to developing LAMP method for each pathogen, we also evaluated the use of a lateral flow device to detect the amplification results and validated the use of dried reagents stable to ambient storage as a step in providing POC LAMP assays.

MATERIALS AND METHODS

LAMP ENZYMES

The discovery and initial characterization of PyroPhage 3173 DNA polymerase and its application in RT-PCR has been described earlier by Schoenfeld et al. (2008) and Moser et al. (2012). The wild type DNA polymerase had a potent proofreading exonuclease activity that was disabled by mutagenesis. The modified enzyme was formulated for use in LAMP and RT LAMP and is commercially available as OmniAmp polymerase (Lucigen Corporation, Middleton, WI). *Bst* DNA polymerase (Lucigen Corporation, WI) was used to compare DNA LAMP assay results with OmniAmp polymerase.

PATHOGENS

Table 1 lists the pathogens for which LAMP assays were developed. All pathogens were obtained from different sources and nucleic acids (DNA or RNA) were extracted from overnight grown cultures (for bacteria) or from cell culture supernatants (for viruses) using commercial kits (Qiagen, Valencia, CA). For Ebola virus (EBoV) and Crimean-Congo hemorrhagic virus (CCHFV), agents of viral hemorrhagic fever, RNA was extracted

Table 1 | List of targets for which LAMP assays were developed using OmniAmp polymerase.

Pathogen	Genome	Target gene	Incubation temp. (°C)	Source
MS2 phage (MS-2)	RNA	Replicase protein (MS2g4)	70	ATCC
Swine influenza virus (SIV) H1N1	RNA	Matrix (M)	72	University of Minnesota, St. Paul, MN
Porcine circovirus-2 (PCV-2)	DNA	Capsid protein (ORF 2)	70	
West Nile virus (WNV)—NY 2001-6263	RNA	Envelope glycoprotein	72	ZeptoMetrix, Buffalo, NY
<i>Edwardsiella ictaluri</i>	DNA	Repetitive element	70	Auburn University, AL
<i>Bacillus atropheus</i> (BAT)	DNA	ATP synthase, β-subunit	68	Steris Corporation, OH
<i>Staphylococcus aureus</i> MSSA	DNA	Carbamate kinase (arcC)	68	ZeptoMetrix, Buffalo, NY
*Ebola virus (EBoV)—Zaire	RNA	Glycoprotein (GP)	72	Galveston National Lab, TX (RNA only)
*Crimean-Congo hemorrhagic fever virus (CCHFV)	RNA	Nucleoprotein (S)	68	
Bovine viral diarrhea virus (BVDV)—type I	RNA	5'-UTR	70	Wisconsin Veterinary Diagnostic Laboratory, Madison, WI

*RNA extracts were provided by Galveston National Laboratory, TX and were certified for use in BSL II facility.

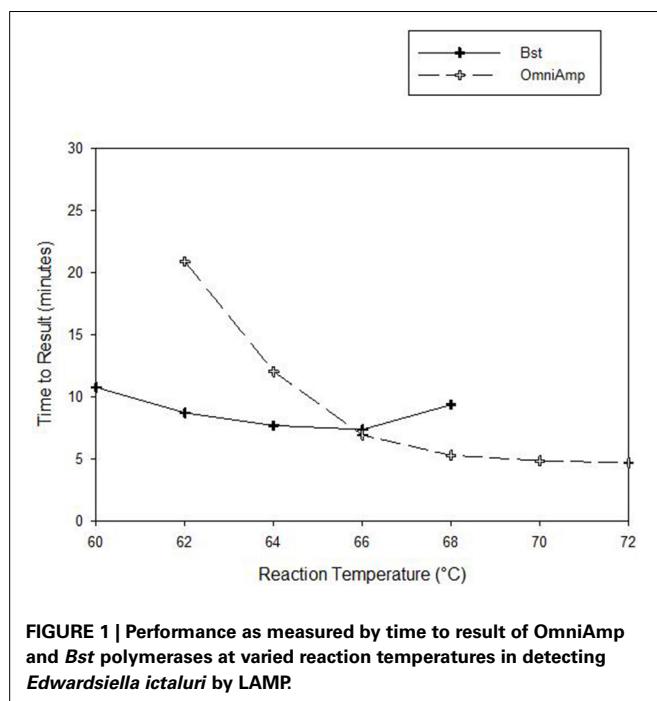
Table 2 | List of LAMP primers used in this report.

Target	Primer	Sequence (5'-3')	Reference
MS2 phage (MS-2)	F3	TGTATGGGATCCGGATGTT	This study
	B3	CAATAGAGCCGCTCTCAGAG	
	FL	CCAGAGAGGAGGTGCCAA	
	BL	TGCAGGATGCAGCGCCTTA	
	FIP	GCCAAACAAACGACGATCGTAAAACCAGCATCCGTAGCCT	
	BIP	GCACGTTCTCCAACGGTCTGGTTGCTTGTACAGCGAAC	
Swine influenza virus (SIV) H1N1	F3	ATCATCCGTCAGGCCCTCA	This study
	B3	TACGCTGCAGTCCTCGCTACTGG	
	FL	TGTTTGCAAGGAAGAAC	
	BL	TCTGACTAAGGGAAATTAGGAT	
	FIP	CCATGAGAGCCTAAGATCAAGCCGAGATCGCACAGA	
	BIP	GACAAGACCAATCCTGTCACCACGGTGAGCGTAACACAA	
Porcine circovirus-2 (PCV-2)	F3	CACTCGTAATGGTTTTATTATTA	Zhao et al., 2011
	B3	TCCACTATTGATTACTCCAAC	
	FL	AACCATGTATGTACAATTCAAGAGAAATTAAATC	
	BL	TTCCAGCAGTTGTAGTCAGC	
	FIP	CAGGAATAACAATCCGTGTAACCATTGGTTAAGTGGGGGTCTT	
	BIP	GAGGCCTACGTGGTCTACATTCAAACACAAAAGAAATCAGCTATG	
West Nile virus (WNV)—NY 2001-6263	F3	TGGATTGGTTCTCGAAGG	This study
	B3	GGTCAGCACGTTGTCATT	
	FL	CATCGATGGTAGGCTTGC	
	BL	TCTCCACCAAAGCTCGT	
	FIP	TTGGCCGCCTCCATATTCACTATTTCAGCTGCCTGACTATCATGT	
	BIP	TGCTATTGGTACCGTCAGCGTTTGAGCTTCTCCATGGTCG	
<i>Edwardsiella ictaluri</i>	F3	CGCGAAAATCATCCCCCT	This study
	B3	ACCCGACAGACAGAGGAAAG	
	FL	GGCAAGAGAGGACGACCACGATA	
	BL	CAGAGACAAGCACGGCGAGTG	
	FIP	ATTGTTGGATGCCCTCCGGGCTGCCTGAGCTTGTCA	
	BIP	TCGAGTCATGGCATTGGCTCCGACACATAGTGGTGGAACG	
<i>Bacillus atrophaeus</i> (BAT)	F3	GTCGCCTAAATGAAGTGC	This study
	B3	GGATAGCGATGAAGAAAGGAC	
	FL	AGGTGAAATGAAGTAGG	
	BL	AAAACGTACGTCAACGA	
	FIP	CGATTAAGTTACAACCAGCAACGACCTCTAGCGTTAACATCGA	
	BIP	ATTCAGGTAAAGTGACCGTCTCCGTTAGCCAAGTAATGGGAC	
<i>Staphylococcus aureus</i> MSSA	F3	TCGAACAGTGACACAACG	This study
	B3	TCTTCTTCGTATAAAAAGGAC	
	FL	CCTATCATAACCCCTGTGACATT	
	BL	ACACGTGTGAAAGTAGATAA	
	FIP	GCGATTGATTCACTTCAACCCATTGGACTTGTGGTC	
	BIP	AGTGTAGAACTGTAGGCACAACTGTTATCAAATCGTGGATCATCT	
Ebola virus—Zaire	F3	ATGGGCTGAAAATGCTACA	This study
	B3	CAGCGAAAGTCGTTCTCG	
	FL	GTCTGGCGCTGCTGGTAGAC	
	BL	CCTTCCACAAAGAGGGTG	
	FIP	TTGTGCACATACCGGCACCGAAAAACCTGACGGGAGTGA	
	BIP	GACCGTGTGCCGGAGACTTGTGGAAGCAAGTCGATCAT	

(Continued)

Table 2 | Continued

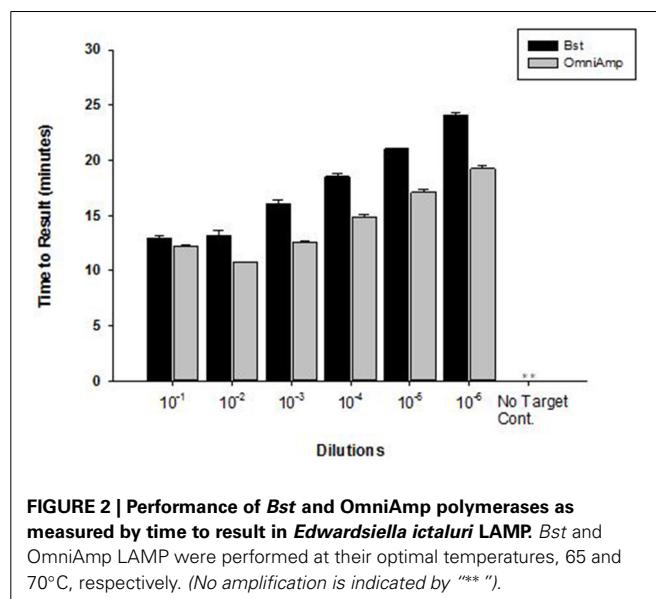
Target	Primer	Sequence (5'-3')	Reference
Crimean-Congo hemorrhagic fever virus (CCHFV)	F3	AGGTGGTTGAAGAGTTCA	This study
	B3	ACAAAACCTTGTGCCCTCC	
	FL	ATAGGAGTTGTGAAGGTGT	
	BL	CCGATGATGCACAGAAGG	
	FIP	TGGGAACACTCTCGAAAAGGAAAAAGGAAATGGACTGTGG	
	BIP	TGTGTTTCAGATGGCCAGTGCGAGATGCGTAGATGGAG	
Bovine viral diarrhea virus (BVDV)—type I	F3	GCGAAGGCCGAAIAGAGG	Koelbl et al., in preparation
	B3	TITGGGCITGCCCTCG	
	BL	CAGGGTAGTCGTCAGTGGTC	
	FIP	CICCCTGCTACCCICCTAACCATGCCCTAGTAGG	
	BIP	CGTTGGATGGCTIAAGCCCTGAGTCCACITGGCATCTCG	

**FIGURE 1 | Performance as measured by time to result of OmniAmp and Bst polymerases at varied reaction temperatures in detecting *Edwardsiella ictaluri* by LAMP.**

in a BSL-4 facility at Galveston National Laboratory, TX and tested for safety for use in BSL-II laboratory before being transferred to Lucigen.

LAMP PRIMER DESIGN

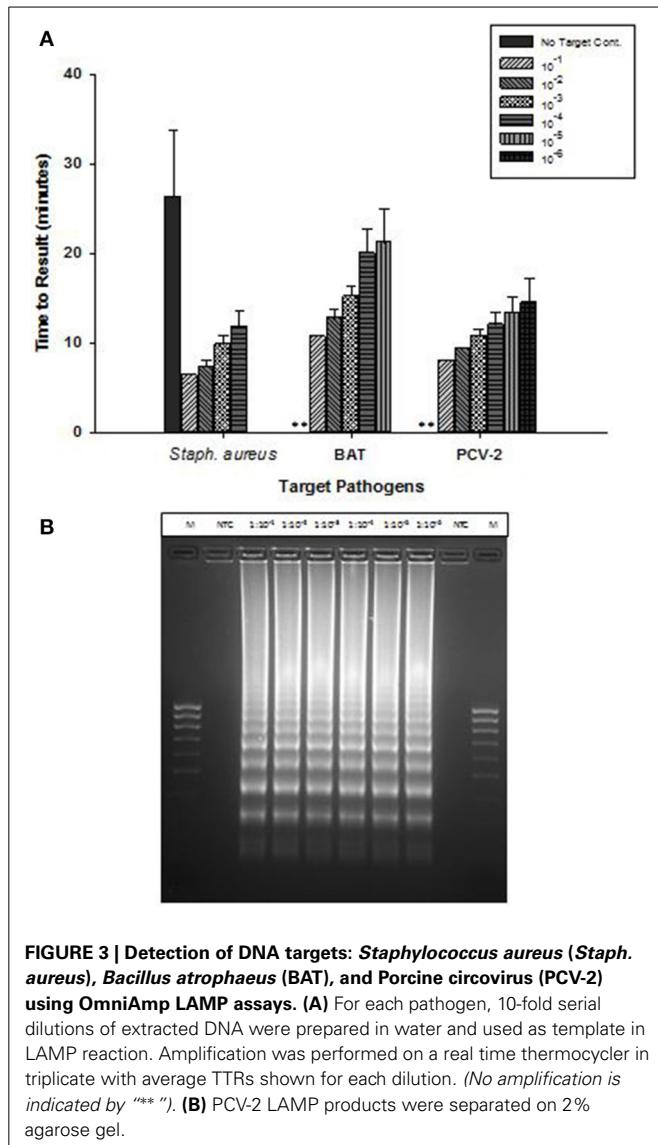
For each pathogen, LAMP primers targeting conserved regions of the indicated pathogens were designed using the online primer design utility, Primer Explorer (<https://primerexplorer.jp/e/>). Conserved regions for the targeted genes were identified by aligning the nucleotide sequences of target genes from GenBank (www.ncbi.nlm.nih.gov) together using clustal W (www.megasoftware.net). Nucleotide sequences (200–300 bp) of the conserved regions as determined by alignment were used to design LAMP primers. Primer designs were selected to provide 100% specificity based on analysis by BLAST (www.ncbi.nlm.nih.gov) and the list of primers is provided in **Table 2**.

**FIGURE 2 | Performance of Bst and OmniAmp polymerases as measured by time to result in *Edwardsiella ictaluri* LAMP. Bst and OmniAmp LAMP were performed at their optimal temperatures, 65 and 70°C, respectively. (No amplification is indicated by “***”).**

For use in LAMP reaction, 20X primer mix was prepared by mixing all six primers (F3,B3:FL,BL:FIP,BIP) in 1:4:8 ratio (Nagamine et al., 2002). Primer mix was stored at -20°C till used.

OPTIMIZATION OF LAMP ASSAY

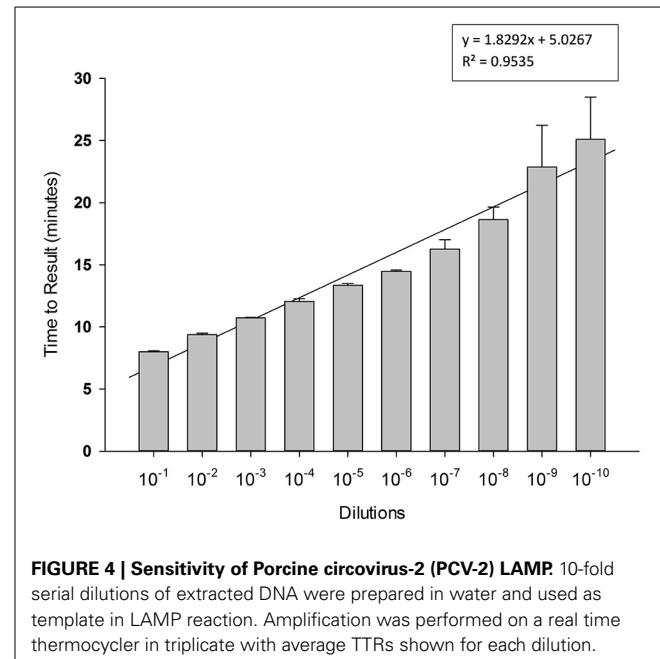
LAMP assays were developed using OmniAmp 2X Isothermal Master Mix (Lucigen Corporation, WI). This master mix is formulated for LAMP and contains optimal concentrations of betaine, salts, dNTPs, and OmniAmp polymerase. Reactions were formulated and performed as described in Lucigen's OmniAmp manual. Final concentration of the reaction mixes were: 1X OmniAmp Master Mix, 2 mM Fiona Green dye (Marker Gene, OR), and 1X LAMP primer mix (IDT, IA; stock solution: 20X); 5 µl of target (DNA or RNA), brought to volume (25 µl) with DNase-RNase free water and incubated in a real time thermocycler (iQ5, Bio-Rad, CA) at constant temperature for indicated times and monitored by detection of Fiona Green fluorescence, measured and quantified by the instrument software at 30 s intervals. The TTR (time to result) was set as the time at which



the fluorescence crossed a hypothetical threshold of 10% of maximal fluorescence. Samples were considered negative if they failed to cross the threshold. In each case, at least three primer sets were synthesized and compared for TTR and specificity. Post-amplification melt analysis was used to distinguish correct (target-dependent) from spurious (target-independent) amplification products. To further verify specificity, reaction products were also visualized by electrophoresis on ethidium bromide-stained 2% agarose gels. Optimal amplification temperatures for each assay were determined using a temperature gradient ranging from 66 to 74°C. To determine the sensitivity of assay, 10-fold serial dilutions of DNA or RNA was prepared in water for detection by LAMP.

DEVELOPMENT OF RAPID SAMPLE PREPARATION METHOD

We also evaluated use of a simple heat lysis method for the extraction of nucleic acid from different clinical matrices. Heat lysis was performed by diluting sample into an extraction buffer followed



by incubation at 90°C for 5 min. After incubation, lysates were used as template in LAMP reaction as described in above section Optimization of LAMP Assay.

For this, sheep whole blood (Hemostat Laboratories, CA) was spiked with *E. coli* MS2 RNA virus particles followed by 10-fold serial dilutions in the same matrix. As a control, 10-fold dilutions of virus particles were made in Tris buffer. Spiked samples were divided into two parts, one part was extracted using a heat lysis method and the other part was used for viral nucleic acid extraction using a commercial kit (QIAamp Viral RNA extraction kit, Qiagen, CA). For heat lysis, samples were diluted in a Tris-EDTA extraction buffer (Lucigen Corporation, WI) and incubated at 90°C for 5 min. After extraction, lysates from both methods were used directly as template in LAMP.

LAMP WITH LYOPHILIZED REAGENTS

To allow ambient storage of formulated LAMP reagents, 1X isothermal master mix, including OmniAmp polymerase was prepared without glycerol, primers, and Fiona green dye. LAMP formulation was lyophilized using BioLyph's (Hopkins, MN) patented technology. Lyophilized LAMP reactions were rehydrated with template, primers and dye into a total volume of 25 µl and incubated and detected in a real time thermocycler isothermally as described above.

DETECTION OF AMPLIFICATION BY USING LATERAL FLOW DEVICE

To simplify detection of positive reactions, we evaluated use of LFD. For this application, forward and reverse loop primers were synthesized with a 5'-conjugated biotin and FITC, respectively. The LF strips were prepared in-house (Lucigen Corporation, Middleton, WI) using an anti-biotin antibody (Thermo Scientific, IL) for capture and a colloidal gold-conjugated anti-FITC antibody (British Biocell International, UK) for detection. In this application LAMP was performed as

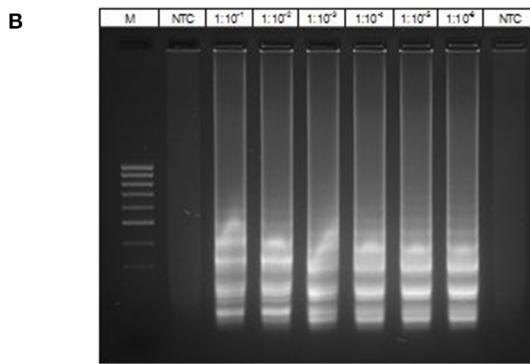
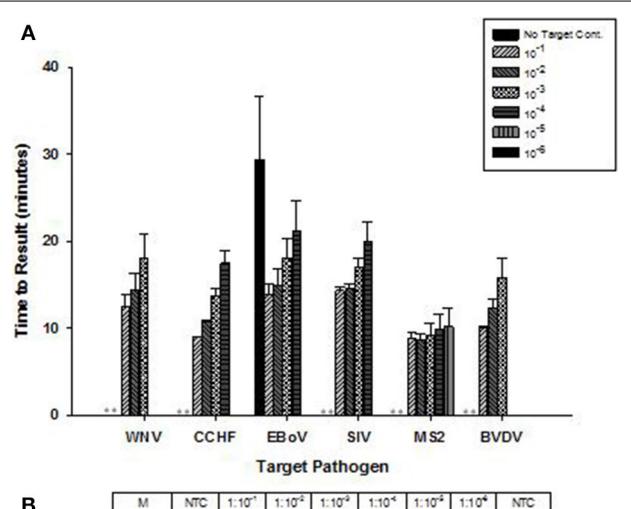


FIGURE 5 | RT-LAMP using OmniAmp polymerase for detection of RNA targets: West Nile virus (WNV), Crimean-Congo hemorrhagic virus (CCHF), Ebola virus (EBoV), Swine influenza virus (SIV), MS2, and Bovine viral diarrhea virus (BVDV). (A) For each pathogen, 10-fold serial dilutions of extracted RNA were prepared in water and used as template in LAMP reaction. Results are averages of three TTR values for each dilution. (No amplification is indicated by “**”). (B) MS2 LAMP products were separated on 2% agarose gel.

described above using labeled loop primers and after completion, reaction products were loaded on the LFD for detection. A positive reaction was indicated by the appearance of red lines at both “Control” and “Test” whereas the appearance of a red line only at “Control” indicates a negative reaction.

This method was evaluated using two strains of *Edwardsiella ictaluri* (S97-9773 and 219). Specificity was determined using one strain each of *Edwardsiella tarda* and *Escherichia coli* (DH10B). For LAMP, six 100-fold dilutions (-2, -4, -6, -8, -10, and -12) of each strain were made in Tryptic Soy Broth (TSB) from overnight grown cultures. These dilutions were used directly as template in *E. ictaluri* LAMP assays. After incubation, LAMP reaction products were loaded on to LFD for visualization.

RESULTS

COMPARISON OF OmniAmp AND *Bst* POLYMERASES

Performance of OmniAmp polymerase in a LAMP reaction was compared with that of *Bst* polymerase (Figure 1). The temperature optimum of OmniAmp is about 70°C, while that of

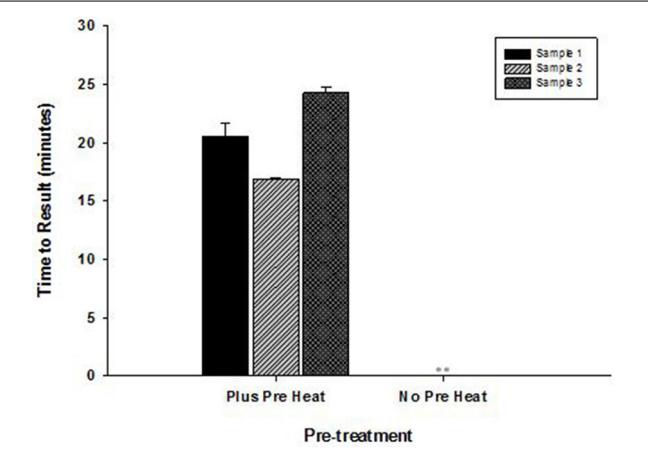


FIGURE 6 | Effect of pre-incubation step on performance of RT-LAMP for detection of Bovine viral diarrhea virus (BVDV)—type 1. RNA extracted from three clinical samples (ear notch) were tested in LAMP with pre-heat step (3 s at 92°C) and without pre-heat step before isothermal incubation at 70°C for 30 min. (No amplification is indicated by “**”).

Bst is 65°C. At its optimal temperature, the OmniAmp polymerase was significantly faster than *Bst* polymerase. This translates to a shorter time to result (TTR), as shown in the detection of the DNA target in *Edwardsiella ictaluri*, an important catfish pathogen. This advantage in shorter TTR was more pronounced at lower template concentrations where detection by the OmniAmp polymerase was 20% faster (Figure 2).

DETECTION OF DNA TARGETS

OmniAmp Pol-based LAMP assays were developed for detection of *Staphylococcus aureus*, *Bacillus atrophaeus* (BAT), and Porcine circovirus (PCV-2), all of which are DNA targets. LAMP primer designs were tested with serial 10-fold dilutions of DNA under optimized reaction conditions. Overall, amplification of all three pathogens was achieved in <30 min with minimal non-specific amplification (Figure 3A). In *S. aureus* LAMP assay, amplification was observed in no-template control (NTC) which was found to be non-specific as it had different melt temperature; melt temperature of specific product: 82.5°C and melt temperature of non-specific product: 84°C (data not shown).

Post-amplification, reaction products were separated on 2% agarose gel and the appearance of ladder like patterns confirmed the correct amplification products. Figure 3B shows PCV-2 LAMP reaction products on 2% gel.

To determine limit of detection, 10-fold serial dilutions of PCV-2 DNA were prepared in water and each dilution was tested in triplicate in LAMP reaction. Results presented in Figure 4 shows high sensitivity of LAMP assay for detection of PCV-2 with limit of detection of about 4 copies of DNA μ l. Regression analysis showed good correlation ($R^2 = 0.95$) between dilutions and time to result (minutes). No amplification signal was detected in any of the negative (no template) controls.

DETECTION OF RNA TARGETS

OmniAmp Pol has inherent reverse transcriptase activity which enables RT-LAMP detection of RNA targets without modification

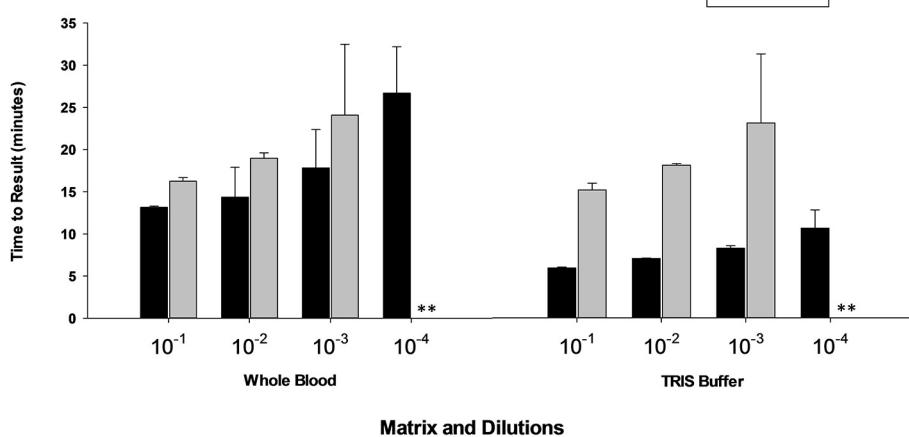


FIGURE 7 | Feasibility of rapid heat lysis method for extraction of viral (MS2) nucleic acid from blood samples. Viral RNA from spiked samples was extracted by two methods: rapid heat lysis (Lucigen) and using a commercial RNA extraction kit (QIAamp, Qiagen). (No amplification is indicated by “**”).

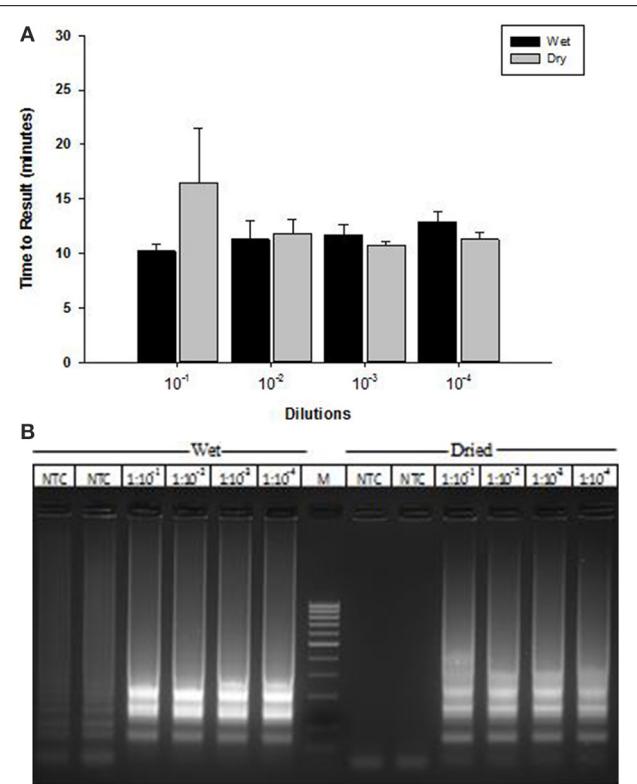


FIGURE 8 | Comparison of MS2 RT-LAMP with wet and dried reagents. (A) Ten-fold serial dilutions of MS2 phage particles was made in water and used directly as template in reaction mixture. Amplification was performed on a real time thermocycler in duplicate and time to results was recorded. (B) LAMP reaction products were separated on 2% agarose gel.

of the formulation used for DNA targets. RT-LAMP assays were developed for six viruses with RNA genomes (Table 1). Results indicate detection in less than 30 min with no additional steps or modification of the reaction formulations (Figure 5A). After

LAMP, reaction products were separated on 2% agarose gel and appearance of ladder like patterns confirmed the correct amplification product as shown in Figure 5B for MS2.

High temperature optimum for OmniAmp was found to be indispensable in detecting certain highly structured regions of genomes especially common in RNA viruses. This advantage of OmniAmp was evaluated in developing a LAMP method for detection of BVDV type 1. Primers designed to target the 5'-UTR failed to amplify when used with standard isothermal incubation at 70°C (not shown). A brief incubation at 92°C for 3 s immediately prior to isothermal incubation allowed efficient amplification (Figure 6).

RAPID SAMPLE PREPARATION METHOD

Results presented in Figure 7 shows that performance of heat lysis method in terms of sensitivity was equivalent to the commercial nucleic acid extraction kit. Presence of inhibitors in whole blood had no effect on the performance of OmniAmp Pol although it did increase TTR when compared to control (dilutions in Tris buffer). We tested the same protocol with other matrices (serum and feces) and results were comparable (data not shown).

LAMP WITH LYOPHILIZED REAGENTS

A lyophilized formulation for the LAMP reagents was compared with wet reagents for detection of MS2 RNA phage target. No diminution of TTR or sensitivity was seen with the dried formulation compared to wet (Figure 8A) nor was an increase in non-specific amplification seen by visualization of LAMP products on a 2% agarose gel (Figure 8B). To evaluate stability, the lyophilized reaction mix (Figure 9A) was incubated at 23, 37, and 45°C and assayed in a *Clostridium difficile* LAMP reaction compared to wet reagent stored at -20°C (Figure 9B). The dried LAMP formulation was stable at 23°C and 37°C for 180 days. The dried reagent was stable at 45°C for 50 days although it did show a measurable drop in TTR after 90 days.

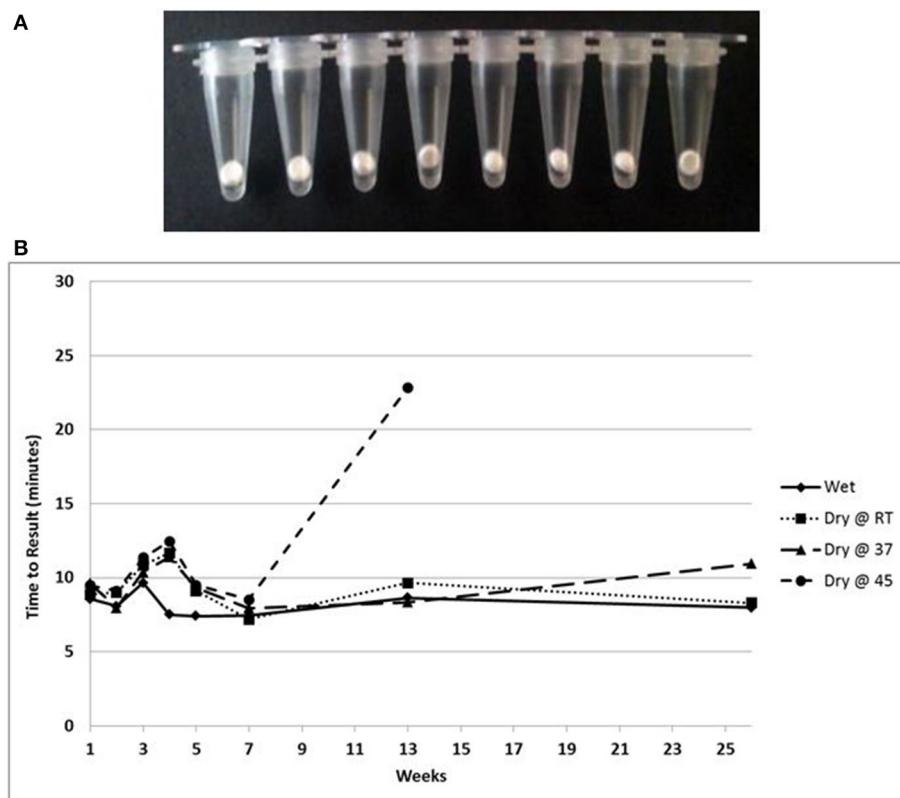


FIGURE 9 | (A) Beads of dried LAMP reagents in PCR tubes. **(B)** Feasibility data showing stability based on TTR of LAMP reagents in dried format stored at the indicated temperatures compared to wet enzyme stored at -20° over 180 days.

LATERAL FLOW DETECTION OF LAMP PRODUCTS

A combination of LAMP and LFD was used to detect the catfish pathogen *E. ictaluri*. For this assay, the LAMP reaction was formulated and performed using the standard protocol except to facilitate LF detection, forward and reverse loop primers were labeled as described in the methods section. Results presented in Figure 10 shows sensitivity and specificity of *E. ictaluri* LAMP as determined using LFD. Appearance of a red line only at “Control” indicated that no product was amplified/or detected in any of the dilutions from both *E. coli* and *E. tarda* cell cultures, confirming 100% specificity of *E. ictaluri* LAMP (Figure 10; lower panel).

In contrast, product was amplified and detected from the cell cultures of both *E. ictaluri* strains as well as positive controls indicating presence of target gene in the samples (Figure 10, upper panel). Positive reaction was detected only in dilutions -2 , -4 , and -6 indicating sensitivity of *E. ictaluri* LAMP equivalent to approximately 8 cells (starting concentration = 10^9 CFU mL $^{-1}$).

DISCUSSION

Nucleic acid tests (NATs) offer major advantages in terms of speed and sensitivity for pathogen detection, but these assays are not simple or inexpensive enough to implement in resource-limited settings. However, development of LAMP technology has changed this paradigm and has given new impetus toward diagnostic methods suitable for use without extensive training or

equipment. LAMP (Notomi et al., 2000; Mori and Notomi, 2009; Mori et al., 2013) is a nucleic acid amplification method that is highly amenable to isothermal detection and best suited to overcome some of the disadvantages of other NATs (PCR, real time PCR). This method uses four or more primers, two of which are engineered to generate loop structures in the nascent strand that primes a cascade of DNA synthesis resulting in microgram yields of amplification product from as low as single-copy targets in as little as 10 min. Well-designed LAMP tests rival real-time PCR in sensitivity and specificity and excel in simplicity of set-up and time to result.

A POC molecular diagnostic test using LAMP based assays is readily achievable, as the only instrument requirement is an inexpensive heater. Portable, battery operated heaters can be improvised (Hernandez et al., 2011) for remote detection amenable to use by individuals with very little training. In some cases, these assays are miniaturized and coupled to hand held devices which would allow instantaneous reporting of results to a central database from virtually any corner of the planet (Stedtfeld et al., 2012; Myers et al., 2013). In other cases, field operation is facilitated by detection of the amplification product using an inexpensive lateral flow device that provides an unambiguous easily interpreted result (Ge et al., 2013). During the last 10 years, LAMP based methods have been developed for detection of various pathogens (Parida et al., 2008; Fu et al., 2011; Mori et al.,

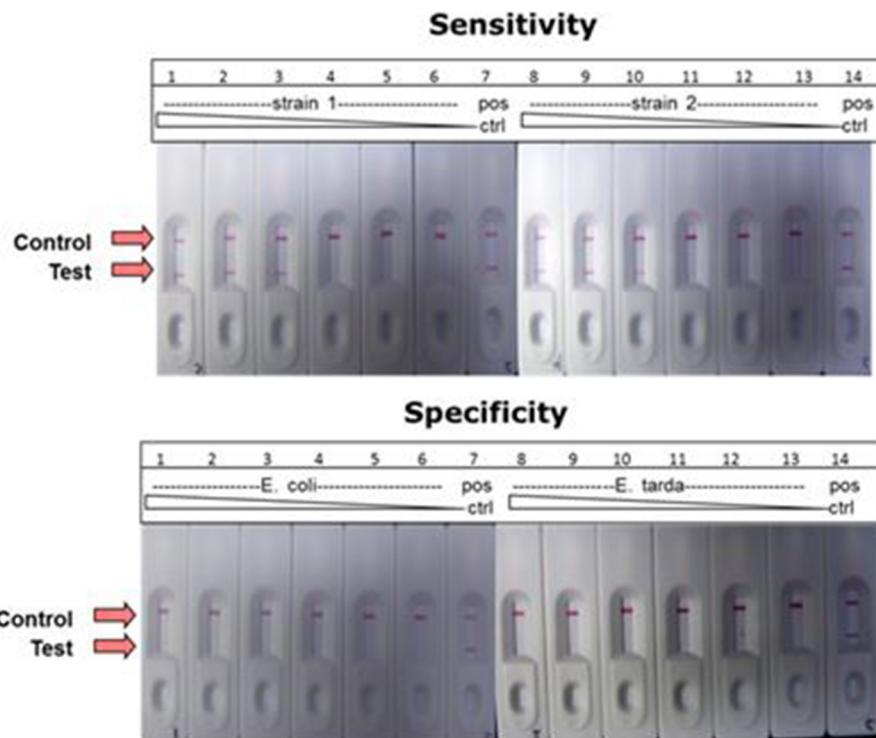


FIGURE 10 | Detection of *E. ictaluri* LAMP reactions by LFD. Sensitivity was determined by testing 100-fold dilutions of *E. ictaluri* S97-9773 (strain 1) and *E. ictaluri* 219 (strain 2) in LAMP (**upper panel**). Dilutions over the same range of *E. coli* and *E. tarda* were tested using for the

same reactions to confirm specificity (**lower panel**). Positive reaction is indicated by the appearance of two red lines, one at “Control” and other at “Test.” Negative reaction is indicated by appearance of only one red line at “Control.”

2013). Conventionally, LAMP uses *Bst* polymerase for amplification of DNA targets. In this paper, we report on applications and advantages of using OmniAmp polymerase in DNA and RT-LAMP reactions.

OmniAmp Pol has a unique combination of properties, including strand displacement, thermostability and reverse transcriptase activity that make it uniquely suitable for use in LAMP formulations for detection of both DNA and RNA targets without modification of the buffer formulation or work flow. In this study, we showed the ability of a single formulation of OmniAmp polymerase to amplify 4 bacterial and viral DNA targets such as *E. ictaluri*, *S. aureus*, *B. atropaeus*, and PCV-2; and 6 RNA viral targets such as WNV, CCHF, EBoV, SIV, MS2, and BVDV. All of the targets amplified in under 30 min with high sensitivity and no alteration of formulation or process. In comparison, RT-LAMP using *Bst* polymerase requires pre-incubation with a reverse transcriptase, typically AMV RT for detection of RNA targets (Notomi et al., 2000; Tanner and Evans, 2014).

Post-incubation, separation of reaction products on 2% agarose gel showed ladder like patterns, which is typical of LAMP (Notomi et al., 2000). In certain cases, where non-specific amplification was observed, melt analysis was used to differentiate between specific and non-specific products. Yamamura et al. (2009) has shown the utility of melt analysis in enabling identification of correct amplification products.

The thermostability of OmniAmp polymerase compared to *Bst* polymerase translates into faster TTR, particularly with more dilute templates. Thermostability is especially important for amplification of GC rich targets or those with extensive secondary structure as high temperature incubation can be used to relax secondary structure. We showed utility of this approach in LAMP method for BVDV. Design parameters for bovine viral diarrhea virus (BVDV) type I LAMP primers is highly constrained by the overall variability of the BVDV genome (Deng and Brock, 1993). This variability limits primer designs to the conserved 5'-UTR, which is highly structured. However, brief incubation at 92°C for 3 s before isothermal incubation enabled amplification through the secondary structure in the 5'-UTR region. In contrast, *Bst* polymerase is not stable above approximately 68°C, and cannot be used for high temperature denaturation of structured targets.

Having a simple and easy to use sample preparation method is one of the major criteria for a true POC diagnostic test. Toward this end, we developed a simple heat lysis method for extraction of nucleic acid and crude lysates used as template in LAMP reaction. No inhibitory effects were observed, indicating that performance of OmniAmp Pol is not impacted by presence of sample matrix components that act as contaminants in PCR based amplification. Another major unmet requirement for POC diagnostics in resource limited settings is long shelf life without maintaining refrigeration or other means of a cold chain (Mabey et al., 2004; Nijru, 2012). The dried formulations described in this report were

stable at ambient temperature (23°C) and 37°C for at least 6 months with no apparent loss in activity.

Positive LAMP reactions can be detected by agarose gel electrophoresis or spectrophotometric measurement of turbidity; however, these methods are not amenable to POC use. Fluorescent detection using dyes such as calcein (Tomita et al., 2008), or HNB (Goto et al., 2009), offers easy to use detection methods. Because these dyes can bind to any dsDNA, they fail to distinguish between specific and non-specific amplification products (Njiru, 2012; Ge et al., 2013). Use of SYBR green I is also not suitable for field applications as it has to be added after the completion of the reaction, a step which increases risk of contamination (Njiru, 2012). LFD has been explored as a means of detecting positive LAMP reactions (Njiru, 2011; Ge et al., 2013; Roskos et al., 2013). In this study, we evaluated LFD in combination with OmniAmp polymerase-based LAMP to visualize amplification products. This method improves specificity due to the secondary binding and detection of amplicon specific targets and negates the need for techniques and instruments unavailable in many low resource settings. The method of labeling the loop primers with biotin and FITC was found to provide high sensitivity and specificity for detection of true positive amplification products. In the present study, we could detect as little as eight cells from two different strains of *E. ictaluri* with no amplification of non-specific targets (*E. tarda* and *E. coli*). These results suggest high sensitivity and specificity of the detection method (LAMP coupled with LF) and shows utility of LFD as a simple and easy to use read out method for visualization of LAMP results.

CONCLUSION

Results presented in this paper show the utility of OmniAmp polymerase in LAMP assays for detecting both RNA and DNA targets. This formulation provides advantages in sample preparation, speed, shelf-stability, and reliability on structured templates compared to traditional LAMP enzymes. We also provide a POC compatible means of detecting positive reactions using LFD.

AUTHOR CONTRIBUTORS

Yogesh Chander helped conceive project, designed and performed experiments and wrote the manuscript. Jim Koelbl, Michael J. Moser, Audrey J. Klingele, Abel Carrias, and Jamie Puckett designed and executed experiments. Mark R. Liles conceived and interpreted experiments. David A. Mead helped conceive project and edit the manuscript. Thomas W. Schoenfeld conceived the project and edited the manuscript.

ACKNOWLEDGMENTS

Authors acknowledge funding received from various agencies: NIH, NSF, and USDA. We thank Drs. Thomas Geisbert and Dennis Bente at Galveston National Laboratory, Galveston, TX for providing VHF RNA extracts.

REFERENCES

- Deng, R., and Brock, V. (1993). 5' and 3' untranslated regions of pestivirus genome: primary and secondary structure analysis. *Nucleic Acid Res.* 21, 1949–1957. doi: 10.1093/nar/21.8.1949
- Fu, S., Qu, G., Guo, S., Ma, L., Zhang, N., Zhang, S., et al. (2011). Applications of loop-mediated isothermal DNA amplification. *Appl. Biochem. Biotechnol.* 163, 845–850. doi: 10.1007/s12010-010-9088-8
- Ge, Y., Wu, B., Qi, X., Zhao, K., Guo, X., Zhu, Y., et al. (2013). Rapid and sensitive detection of novel avian-origin influenza A (H7N9) virus by reverse transcription loop-mediated isothermal amplification combined with a lateral-flow device. *PLoS ONE* 8:e69941. doi: 10.1371/journal.pone.0069941
- Goto, M., Honda, E., Ogura, A., Nomoto, A., and Hanaki, K. I. (2009). Colorimetric detection of loop-mediated isothermal amplification reaction by using hydroxy naphthol blue. *Biotechniques* 46, 167–172. doi: 10.2144/000113072
- Hernandez, L. O., Negussie, H., Mazzola, L. T., Laser, D. J., and Droitcour, A. (2011). Integrated heaters for temperature control in disposable bioassay cartridges for use with portable, battery-operated instruments. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 326–332. doi: 10.1109/IEMBS.2011.6090085
- Huang, W. H., Zhai, F., and Hong, G. F. (1999). Cloning, expression, characterization and application of Bst DNA polymerase large fragment. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* 31, 379–384.
- Mabey, D., Peeling, R. W., Ustianowski, A., and Perkins, M. D. (2004). Diagnostics for the developing world. *Nat. Rev. Microbiol.* 2, 231–240. doi: 10.1038/nrmicro841
- Mori, Y., Kanda, H., and Notomi, T. (2013). Loop-mediated isothermal amplification (LAMP): recent progress in research and development. *J. Infect. Chemother.* 19, 404–411. doi: 10.1007/s10156-013-0590-0
- Mori, Y., and Notomi, T. (2009). Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. *J. Infect. Chemother.* 15, 62–69. doi: 10.1007/s10156-009-0669-9
- Moser, M. J., Difrancesco, R. A., Gowda, K., Klingele, A. J., Sugar, D. R., Stocki, S., et al. (2012). Thermostable DNA polymerase from a viral metagenome is a potent rt-PCR enzyme. *PLoS ONE* 7:e38371. doi: 10.1371/journal.pone.0038371
- Myers, F. B., Henrikson, R. H., Bone, J., and Lee, L. P. (2013). A handheld point-of-care genomic diagnostic system. *PLoS ONE* 8:e70266. doi: 10.1371/journal.pone.0070266
- Nagamine, K., Hase, T., and Notomi, T. (2002). Accelerated reaction loop-mediated isothermal amplification using loop primers. *Mol. Cell. Probes* 16, 223–239. doi: 10.1006/mcpr.2002.0415
- Njiru, Z. K. (2012). Loop-mediated isothermal amplification technology: towards point of care diagnostics. *PLoS Negl. Trop. Dis.* 6:e1572. doi: 10.1371/journal.pntd.0001572
- Njiru, Z. K. (2011). Rapid and sensitive detection of human African trypanosomiasis by loop-mediated isothermal amplification combined with a lateral-flow dipstick. *Diagn. Microbiol. Infect. Dis.* 69, 205–209. doi: 10.1016/j.diagmicrobio.2010.08.026
- Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., et al. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* 28:E63. doi: 10.1093/nar/28.12.e63
- Parida, M., Sannarangaiah, S., Dash, P. K., Rao, P. V., and Morita, K. (2008). Loop mediated isothermal amplification (LAMP): a new generation of innovative gene amplification technique; perspectives in clinical diagnosis of infectious diseases. *Rev. Med. Virol.* 18, 407–421. doi: 10.1002/rmv.593
- Roskos, K., Hickerson, A. I., Lu, H. W., Ferguson, T. M., Shinde, D. N., Klaue, Y., et al. (2013). Simple system for isothermal DNA amplification coupled to lateral flow detection. *PLoS ONE* 8:e69355. doi: 10.1371/journal.pone.0069355
- Schoenfeld, T., Patterson, M., Richardson, P. M., Wommack, K. E., Young, M., and Mead, D. (2008). Assembly of viral metagenomes from yellowstone hot springs. *Appl. Environ. Microbiol.* 74, 4164–4174. doi: 10.1128/AEM.02598-07
- Segawa, T., Kobayashi, Y., Sase, Y., Itou, T., Suzuki, M., Endoh, T., et al. (2014). Easy-to-use rapid gene amplification method for detection of RNA and DNA viruses in sera and feces from various animals. *J. Virol. Methods* 201, 31–37. doi: 10.1016/j.jviromet.2014.01.019
- Stedtfeld, R. D., Tourlousse, D. M., Seyrig, G., Stedtfeld, T. M., Kronlein, M., Price, S., et al. (2012). Gene-Z: a device for point of care genetic testing using a smartphone. *Lab Chip* 12, 1454–1462. doi: 10.1039/c2lc21226a
- Tanner, N. A., and Evans, T. C. (2014). Loop-mediated isothermal amplification for detection of nucleic acids. *Curr. Protoc. Mol. Biol.* 105, unit 15.14. doi: 10.1002/0471142727.mb1514s105
- Tomita, N., Mori, Y., Kanda, H., and Notomi, T. (2008). Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nat. Protoc.* 3, 877–882. doi: 10.1038/nprot.2008.57

- Yamamura, M., Makimura, K., and Ota, Y. (2009). Evaluation of a new rapid molecular diagnostic system for *Plasmodium falciparum* combined with DNA filter paper, loop-mediated isothermal amplification, and melting curve analysis. *Jpn. J. Infect. Dis.* 62, 20–25.
- Zhao, K., Shi, W., Han, F., Xu, Y., Zhu, L., Zou, Y., et al. (2011). Specific, simple and rapid detection of porcine circovirus type 2 using the loop-mediated isothermal amplification method. *Virol. J.* 8:126. doi: 10.1186/1743-422X-8-126

Conflict of Interest Statement: Yogesh Chander, Jim Koelbel, Jamie Puckett, Michael Moser, Audrey Klingele, David Mead, and Thomas Schoenfeld are employed by Lucigen Corporation. Lucigen has commercialized the OmniAmp polymerase for research use only. Mark Liles and Abel Carrias have no commercial or financial relationship with Lucigen Corporation, WI and declare no conflict of interest.

Received: 15 April 2014; accepted: 14 July 2014; published online: 01 August 2014.

Citation: Chander Y, Koelbl J, Puckett J, Moser MJ, Klingele AJ, Liles MR, Carrias A, Mead DA and Schoenfeld TW (2014) A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.* 5:395. doi: 10.3389/fmicb.2014.00395

This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Chander, Koelbl, Puckett, Moser, Klingele, Liles, Carrias, Mead and Schoenfeld. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.