

Discovery notes

Open Access

## A highly conserved family of inactivated archaeal B family DNA polymerases

Igor B Rogozin<sup>1</sup>, Kira S Makarova<sup>1</sup>, Youri I Pavlov<sup>2</sup> and Eugene V Koonin\*<sup>1</sup>

Address: <sup>1</sup>National Center for Biotechnology Information NLM, National Institutes of Health, Bethesda, MD, 20894, USA and <sup>2</sup>Eppley Institute for Research in Cancer and Allied Diseases, Departments of Biochemistry and Molecular Biology, and Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, 68198-6805, USA

Email: Igor B Rogozin - rogozin@ncbi.nlm.nih.gov; Kira S Makarova - makarova@ncbi.nlm.nih.gov; Youri I Pavlov - ypavlov@unmc.edu; Eugene V Koonin\* - koonin@ncbi.nlm.nih.gov

\* Corresponding author

Published: 6 August 2008

Received: 25 July 2008

*Biology Direct* 2008, **3**:32 doi:10.1186/1745-6150-3-32

Accepted: 6 August 2008

This article is available from: <http://www.biology-direct.com/content/3/1/32>

© 2008 Rogozin et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

A widespread and highly conserved family of apparently inactivated derivatives of archaeal B-family DNA polymerases is described. Phylogenetic analysis shows that the inactivated forms comprise a distinct clade among archaeal B-family polymerases and that, within this clade, Euryarchaea and Crenarchaea are clearly separated from each other and from a small group of bacterial homologs. These findings are compatible with an ancient duplication of the DNA polymerase gene followed by inactivation and parallel loss in some of the lineages although contribution of horizontal gene transfer cannot be ruled out. The inactivated derivative of the archaeal DNA polymerase could form a complex with the active paralog and play a structural role in DNA replication.

### Reviewers

This article was reviewed by Purificacion Lopez-Garcia and Chris Ponting. For the full reviews, please go to the Reviewers' Reports section.

### Findings

DNA polymerases are enzymes that are essential for genome replication and repair in all cellular life forms [1,2]. There are several distinct families of DNA polymerases some of which are unrelated to each other whereas most show varying degrees of relationship [3,4]. All archaea and eukaryotes encode at least one but typically several paralogous B-family DNA polymerases that play a key role in DNA replication [5]. Euryarchaeota typically possess one or two B-family polymerase, in addition to the apparently unrelated D-family polymerase whereas Cre-

narchaeota have two or three paralogous B-family polymerases but no D-family polymerases [6-8].

When examining the relationships between archaeal B-family DNA polymerases, we unexpectedly observed that in many Crenarchaeota and Euryarchaeota, one of the B-family DNA polymerase paralogs, despite the high level of overall sequence conservation, contains disrupted versions of the sequence motifs that are known to be essential for the catalytic functions. The polymerases of the B-family contain two enzymatic domain, namely, the N-ter-

minimal 3'-exonuclease domain that performs proofreading and the C-terminal polymerase proper domain [9,10]. In the apparently inactivated archaeal polymerase, the sequence motifs that in the active forms harbor the catalytic amino acids required for each of these activities are partially disrupted (Fig. 1 and Additional File 1). In particular, the crucial, most conserved YGDTD motif that coordinates  $Mg^{2+}$  in B-family DNA polymerases and in which both aspartates and the tyrosine are essential for the polymerase activity [11-13] harbors at least two replacements in the apparently inactivated derivatives although it is notable that the distal aspartate is conserved and so is likely to retain an important function (Fig. 1). In addition, the most conserved DIE motif (Exo I) of the N-terminal 3'-exonuclease proofreading domain does not have a clear counterpart in the corresponding part of the polymerase derivatives described here (Additional File 1), strongly suggesting that the exonuclease activity is inactivated as well.

Phylogenetic analysis of archaeal B-family polymerases showed that the inactivated forms comprised a distinct clade with a 100% bootstrap support (Fig. 1). Moreover, within this clade, the Euryarchaeal and Crenarchaeal forms were clearly separated, also with full bootstrap support, and joined a third subclade that included similarly inactivated homologs from three diverse bacteria (Fig. 1). The presence of the inactivated DNA polymerases in diverse subsets of both Euryarchaeota and Crenarchaeota is compatible with an ancient duplication of the DNA polymerase gene followed by inactivation and parallel loss in some of the lineages. However, a relatively later duplication, inactivation and subsequent spread via horizontal gene transfer also could be a viable evolutionary scenario for the inactivated polymerases.

Inactivation of the enzymatic function of a protein and utilization of the inactivated protein for non-enzymatic (structural) roles, so that it retains substantial sequence conservation, is not without precedent. Perhaps, the cases that are most relevant for the present finding are the apparent inactivation of the small subunits of DNA polymerases and the N-terminal 3'-exonuclease domains in eukaryotes. In all Euryarchaeota, the small subunits of PolD contain all sequence signature of the calcineurine-like superfamily of phosphoesterases [14] and have been shown to possess 3'-exonuclease, proofreading activity [15-17]. By contrast, in the eukaryotic orthologs of these proteins, most of the catalytic residues are replaced, so the protein apparently performs a structural role in the polymerase complex [8,14]. Another example is the inactivation of the N-terminal 3'-exonuclease domain in two eukaryotic members of the polB family, pol  $\alpha$  and pol  $\zeta$ . Despite the overall high level of sequence conservation, the catalytic residues in the exonuclease domains of these

polymerases are replaced with amino acid residues that cannot function in the proofreading reaction, so structural roles have been proposed for these domains [2,18]. To our knowledge, the PolB derivative described here is the first case of an evolutionarily conserved and, by implication, functionally important but, apparently, inactive DNA polymerase. Conceivably, this protein is involved in archaeal DNA replication, perhaps, as a structural subunit of the DNA polymerase holoenzyme [14,19]. Determination of the function of this unusual protein should yield significant insight into archaeal DNA replication.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

IBR and YIP made the original observation and performed initial sequence comparisons, KSM performed the final sequence analysis and phylogenetic analysis, EVK wrote the manuscript. All authors edited and approved the final version.

### Reviewers' comments

#### Reviewer's report 1

*Purificacion Lopez-Garcia, Universite Paris-Sud*

Rogozin and co-workers report here an interesting observation. They identify a highly conserved clade of B-family polymerases grouping euryarchaeotal, crenarchaeotal and also some bacterial sequences whose  $Mg^{2+}$ -binding motif has several replacements from the canonical YGDTD, notably the tyrosine and one of the aspartates that are essential during polymerization. They propose that these amino acid changes lead to inactivated enzymes that might have retained a purely structural role. This is a justified in silico prediction that, however, remains to be tested. In particular, it would be important to see whether the enzymatic activity actually disappears or whether some kind of activity can still be performed, or performed under particular conditions and with precise partners. One of the aspartates involved in polymerization remains very well conserved in this motif; the same is true for another aspartate upstream in the sequence. Such sequence conservation at long evolutionary distances might indicate more than an exclusive structural function. I hope that this observation will foster the appropriate biochemical work downstream. I would recommend the authors to insist on the hypothetical nature of their prediction, which they do in parts of the text by speaking about 'apparently inactivated' polymerases, through all of it including the title and the figure legend.

*Authors' response: We appreciate the comments and included additional qualifications in the text; we also emphasize the conservation and probable functional importance of the distal aspartate in the  $Mg^{2+}$ -binding motif. However, we also stress*

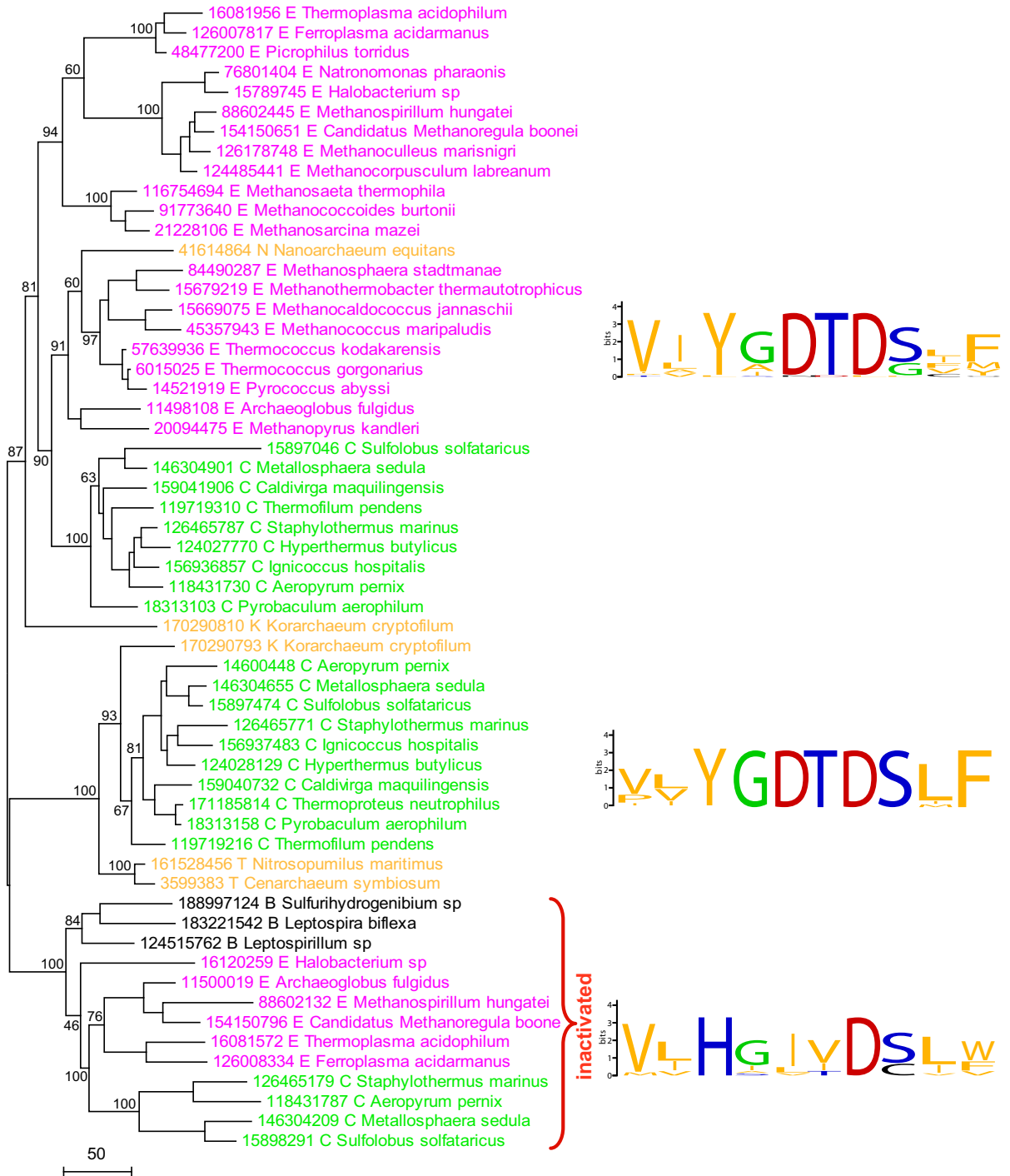


Figure 1

**Figure 1****Inactivated derivatives of B-family DNA polymerases in archaea: inactivation of the principal catalytic motif and phylogeny.**

The right part of the figure shows the sequences of the essential Mg<sup>2+</sup>-binding motif of B-family DNA polymerases for two families of active archaeal polymerase (top and middle) and the inactivated derivative (bottom). The motifs are shown in the form of sequence LOGOs where the height of the amino acid symbols is a function of the frequency of the given amino acid in the given position [20,21]. The left part of the figure shows a phylogenetic tree for archaeal B-family DNA polymerases. Color code: purple, Euryarchaeota (E); green: Crenarchaeota (C); brown: organisms that might represent distinct archaeal phyla: Nanoarchaeota (N), Korarchaeota (K), and Thaumarchaeota (T); black, bacteria (B). Each organism is denoted by the full systematic name and the Gene Identifier (GI) number. Multiple sequence alignment of archaeal B-family polymerases (Additional File 1) was constructed using the MUSCLE program [22]. The tree was built using the maximum likelihood method implemented in the MOLPHY program [23] by local rearrangement of an original Fitch tree [24]. The same program was used to compute bootstrap values which are indicated (%) for selected major branches.

*the obvious disruption of the catalytic motifs in the exonuclease domain, an observation that suggests concordant inactivation of both domains of these proteins.*

**Reviewer's report 2**

Chris Ponting, Oxford University

Rogozin et al. have made the interesting observation that some bacteria, euryarchaea and crenarchaea possess a homologue of B family DNA polymerase that contains substitutions within the YGDTD active site sequence. The evidence that this represents a homologue whose function is likely to be distinct from that of typical family B polymerases is compelling. What function this might be is uncertain, although the authors tentatively suggest an atypical role in DNA replication. This study shows, lest we ever forget, that demonstration of homology need not always lead to a successful prediction of function.

**Additional material****Additional File 1**

Multiple sequence alignment of active and inactivated forms of archaeal B-family DNA polymerases.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1745-6150-3-32-S1.doc>]

**Acknowledgements**

The research of IBR, KSM and EVK is supported by the Department of Health and Human Services intramural program (NIH, National Library of Medicine). YIP is supported by the NIH grant 1R01CA129925-01A1, Nebraska DHHS grant LB506 and Eppley Institute Pilot Project grant.

**References**

- Hubscher U, Maga G, Spadari S: **Eukaryotic DNA polymerases.** *Annu Rev Biochem* 2002, **71**:133-163.
- Pavlov YI, Shcherbakova PV, Rogozin IB: **Roles of DNA polymerases in replication, repair, and recombination in eukaryotes.** *Int Rev Cytol* 2006, **255**:41-132.
- Braithwaite DK, Ito J: **Compilation, alignment, and phylogenetic relationships of DNA polymerases.** *Nucleic Acids Res* 1993, **21**(4):787-802.
- Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, Lawrence CW, Nakanishi M, Ohmori H, Prakash L, Prakash S, Reynaud CA, Sugino A, Todo T, Wang Z, Weill JC, Woodgate R: **Eukaryotic DNA polymerases: proposal for a revised nomenclature.** *J Biol Chem* 2001, **276**(47):43487-43490.
- Cann IK, Ishino Y: **Archaeal DNA replication: identifying the pieces to solve a puzzle.** *Genetics* 1999, **152**(4):1249-1267.
- Cann IK, Komori K, Toh H, Kanai S, Ishino Y: **A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase.** *Proc Natl Acad Sci U S A* 1998, **95**(24):14250-14255.
- Ishino Y, Komori K, Cann IK, Koga Y: **A novel DNA polymerase family found in Archaea.** *J Bacteriol* 1998, **180**(8):2232-2236.
- Gueguen Y, Rolland JL, Lecompte O, Azam P, Le Romancer G, Flament D, Raffin JP, Dietrich J: **Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon Pyrococcus abyssi.** *Eur J Biochem* 2001, **268**(22):5961-5969.
- Zhao Y, Jeruzalmi D, Moarefi I, Leighton L, Lasken R, Kuriyan J: **Crystal structure of an archaeobacterial DNA polymerase.** *Structure* 1999, **7**(10):1189-1199.
- Savino C, Federici L, Johnson KA, Vallone B, Nastopoulos V, Rossi M, Pisani FM, Tsernoglou D: **Insights into DNA replication: the crystal structure of DNA polymerase BI from the archaeon Sulfolobus solfataricus.** *Structure* 2004, **12**(11):2001-2008.
- Bernad A, Blanco L, Salas M: **Site-directed mutagenesis of the YCDTDS amino acid motif of the phi 29 DNA polymerase.** *Gene* 1990, **94**(1):45-51.
- Blasco MA, Bernad A, Blanco L, Salas M: **Characterization and mapping of the pyrophosphorolytic activity of the phage phi 29 DNA polymerase. Involvement of amino acid motifs highly conserved in alpha-like DNA polymerases.** *J Biol Chem* 1991, **266**(12):7904-7909.
- Copeland WC, Wang TS: **Mutational analysis of the human DNA polymerase alpha. The most conserved region in alpha-like DNA polymerases is involved in metal-specific catalysis.** *J Biol Chem* 1993, **268**(15):11028-11040.
- Aravind L, Koonin EV: **Phosphoesterase domains associated with DNA polymerases of diverse origins.** *Nucleic Acids Res* 1998, **26**(16):3746-3752.
- Jokela M, Eskelinen A, Pospiech H, Rouvinen J, Syvaaja JE: **Characterization of the 3' exonuclease subunit DPI of Methanococcus jannaschii replicative DNA polymerase D.** *Nucleic Acids Res* 2004, **32**(8):2430-2440.
- Jokela M, Raki M, Heikkinen K, Sepponen K, Eskelinen A, Syvaaja JE: **The screening of expression and purification conditions for replicative DNA polymerase associated B-subunits, assignment of the exonuclease activity to the C-terminus of archaeal pol D DPI subunit.** *Protein Expr Purif* 2005, **43**(1):73-84.
- Shen Y, Tang XF, Yokoyama H, Matsui E, Matsui I: **A 21-amino acid peptide from the cysteine cluster II of the family D DNA polymerase from Pyrococcus horikoshii stimulates its nucle-**

- ase activity which is Mre I I-like and prefers manganese ion as the cofactor. *Nucleic Acids Res* 2004, **32(1)**:158-168.
18. Garg P, Burgers PM: **DNA polymerases that propagate the eukaryotic DNA replication fork.** *Crit Rev Biochem Mol Biol* 2005, **40(2)**:115-128.
  19. Frouin I, Montecucco A, Spadari S, Maga G: **DNA replication: a complex matter.** *EMBO Rep* 2003, **4(7)**:666-670.
  20. Schneider TD, Stormo GD, Gold L, Ehrenfeucht A: **Information content of binding sites on nucleotide sequences.** *J Mol Biol* 1986, **188(3)**:415-431.
  21. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.
  22. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32(5)**:1792-1797.
  23. Adachi J, Hasegawa M: **MOLPHY: Programs for molecular phylogenetics.** In *Computer Science Monographs 27*; Tokyo, Institute of Statistical Mathematics; 1992.
  24. Felsenstein J: **Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods.** *Methods Enzymol* 1996, **266**:418-427.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

