3746–3752 *Nucleic Acids Research, 1998, Vol. 26, No. 16*

Phosphoesterase domains associated with DNA polymerases of diverse origins

L. Aravind[1,2] and Eugene V. Koonin[2,*]

[1]Department of Biology, Texas A&M University, College Station, TX 77843, USA and [2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received April 20, 1998; Revised and Accepted June 25, 1998

ABSTRACT

Computer analysis of DNA polymerase protein sequences revealed previously unidentified conserved domains that belong to two distinct superfamilies of phosphoesterases. The α subunits of bacterial DNA polymerase III and two distinct family X DNA polymerases are shown to contain an N-terminal domain that defines a novel enzymatic superfamily, designated PHP, after polymerase and histidinol phosphatase. The predicted catalytic site of the PHP superfamily consists of four motifs containing conserved histidine residues that are likely to be involved in metal-dependent catalysis of phosphoester bond hydrolysis. The PHP domain is highly conserved in all bacterial polymerase III α subunits, but in proteobacteria and mycoplasmas, the conserved motifs are distorted, suggesting a loss of the enzymatic activity. Another conserved domain, found in the small subunits of archaeal DNA polymerase II and eukaryotic DNA polymerases α and δ, is shown to belong to the superfamily of calcineurin-like phosphoesterases, which unites a variety of phosphatases and nucleases. The conserved motifs required for phosphoesterase activity are intact in the archaeal DNA polymerase subunits, but are disrupted in their eukaryotic orthologs. A hypothesis is proposed that bacterial and archaeal replicative DNA polymerases possess intrinsic phosphatase activity that hydrolyzes the pyrophosphate released during nucleotide polymerization. As proposed previously, pyrophosphate hydrolysis may be necessary to drive the polymerization reaction forward. The phosphoesterase domains with disrupted catalytic motifs may assume an allosteric, regulatory function and/or bind other subunits of DNA polymerase holoenzymes. In these cases, the pyrophosphate may be hydrolyzed by a stand-alone phosphatase, and candidates for such a role were identified among bacterial PHP superfamily members.

INTRODUCTION

DNA polymerases are central to genome replication and repair (1–3). In spite of their fundamental role, these enzymes are very different among different life forms (4). DNA polymerases have been classified on the basis of sequence similarity between their large subunits or domains that catalyze nucleotide polymerization. There are at least four distinct DNA polymerase superfamilies that bear little, if any, relationship to each other, namely: (i) bacterial DNA polymerase I and its eukaryotic orthologs; (ii) B family DNA polymerases including eukaryotic and archaeal replicative and repair polymerases, polymerases of animal viruses and bacteriophages, and *Escherichia coli* polymerase II; (iii) bacterial replicative DNA polymerase III; and (iv) family X polymerases, which includes terminal nucleotidyltransferases, several polymerases involved in repair, and a variety of other diverse nucleotidyltransferases (1–3 and references therein). Typically, in addition to the subunits or domains directly involved in polymerization, DNA polymerase holoenzymes contain additional subunits (domains) that perform accessory functions in replication or repair (1,2,5).

Here we present the results of detailed computer analysis of the sequences of DNA polymerase subunits. We show that three distinct groups of DNA polymerases, namely the α-subunit of bacterial DNA polymerase III, family X polymerases and the small subunits of archaeal DNA polymerases and eukaryotic polymerases α, δ and ε, contain previously undetected, conserved domains. These domains belong to two distinct, ancient superfamilies of enzymes with phosphatase, and in some cases nuclease, activity. One of these superfamilies, whose only functionally characterized member is yeast histidinol phosphatase, has not been recognized previously. The association of the phosphoesterase domains with DNA polymerases clearly is the result of at least two independent evolutionary events, which suggests an important role for these domains in the polymerase function. We hypothesize that this primary role is the hydrolysis of the released pyrophosphate, which shifts the reaction equilibrium towards nucleotide polymerization.

MATERIALS AND METHODS

Databases and sequence analysis

The non-redundant (NR) protein sequence database at the National Center for Biotechnology Information (NIH, Bethesda) was searched using the gapped BLAST program (6). Database searches were iterated using the PSI-BLAST program, which constructs position-dependent weight matrices on the basis of alignments generated by BLAST and employs them for subsequent

*To whom correspondence should be addressed. Tel: +1 301 453 5913; Fax: +1 301 480 9241; Email: koonin@ncbi.nlm.nih.gov

search iterations (6). In addition to the automatic generation of profiles 'on the fly', a new option of the PSI-BLAST program was used to generate matrices from alignments constructed independently and seed PSI-BLAST iterations with these (A.Schäffer, L.Aravind and E.V.Koonin, unpublished work). Multiple sequence alignments were constructed using the CLUSTALW program (7) or the Gibbs-sampling option of the MACAW program (8,9). The program MoST was used for searching the NR database with motifs represented as ungapped alignment blocks (10). Clustering of sequences by sequence similarity was performed using the CLUS program (11). Protein secondary structure was predicted using the PHD program (12), and the PHD-based threading was used to assess possible relationships between aligned protein sequences and known structural folds (13). Globular domains in proteins were predicted using the SEG program with the following parameters: window length 45, trigger complexity 3.4, extension complexity 3.75 (14).

RESULTS AND DISCUSSION

PHP domain: a novel family of phosphoesterases associated with polIII α subunit and X-family polymerases

DNA polymerase III α subunits from bacteria and the DNA polymerases of the X family from *Methanobacterium thermoautotrophicum*, *Aquifex aeolicus* and *Bacillus subtilis* (but not the homologous DNA polymerase from *Thermus aquaticus*) showed statistically significant sequence similarity to each other when the NR database was searched with any of these sequences as the query using the gapped BLAST program [expectation (e) value: ~10^{-3}–10^{-4}]. This was an unexpected finding given that these DNA polymerases belong to two different superfamilies (3). Examination of the distribution of the detected similarity among the predicted globular domains in the two families of polymerases showed that the C-terminal domain of the *B.subtilis* and the archaeal X family DNA polymerase was related to a domain typically found at the extreme N-terminus of the DNA polymerase III α subunit. This domain was clearly distinct from the well-defined nucleotidyltransferase domain of the X family (15,16) and the proposed functionally equivalent catalytic region of the polIII α subunits (17), which do not contain any detectable common motifs. This novel domain shared by the two diverse families of polymerases was subjected to further iterative searches using the PSI-BLAST program, which was run to convergence. These searches resulted in the delineation of a previously uncharacterized superfamily of domains that occur both as stand-alone proteins in a variety of bacterial and archaeal genomes and as fusions with the polymerase domains (Fig. 1). The internal consistency of the derived protein set was assessed by carrying out multiple PSI-BLAST searches with different members of the superfamily and evaluating recovery of the other members at e-value cut-offs below 10^{-4} within four iterations. This helped to unambiguously define the superfamily as different starting queries retrieved the rest of the superfamily members at a statistically significant level. Additionally, the motif searching program MoST was used to scan the non-redundant database with weighted profiles of the motifs derived from the PolIII α subunit version of the domain with r values of 0.005 and 0.001. This resulted in consistent recovery of the rest of the superfamily without any additions to the set defined by the multiple PSI-BLAST analysis.

As shown by multiple alignment analysis using a combination of PSI-BLAST derived alignments and Gibbs sampling as incorporated in the MACAW program, the polymerase and histidinol phosphatase (PHP) domain described here consists of four conserved core regions (motifs I–IV in Fig. 1). The characteristic feature of these motifs is the presence of highly conserved histidines and aspartates, which are likely to participate in catalysis. Motif I has a dyad of histidines which are separated by a single amino acid. A similar arrangement is also observed in two other groups of metal-dependent hydrolases, namely the urease-pyrimidinase superfamily (18) and the calcineurin-like phosphoesterase superfamily (19; see also below), with the conserved histidines apparently coordinating a metal ion such as zinc or nickel. Motifs II–IV contain additional conserved histidines and aspartates (Fig. 1), which may be involved in catalysis by participating in electron transfer and/or through metal coordination. Motif IV is missing in some archaeal and bacterial proteins (Fig. 1), suggesting the possibility that PHP family proteins may differ in the number of coordinated metal ions. The nature and distribution of the conserved residues in the PHP superfamily is analogous to the urease-pyrimidinase superfamily (18). Thus it seems likely that similarly to the enzymes of this superfamily, the activity of the PHP phosphoesterases requires two coordinated metal cations. However, database searches using profiles derived from multiple alignments of each of these superfamilies failed to detect any similarity between them, and furthermore, the distance between motifs I and II in the PHP superfamily is strictly constrained and is much shorter than the distance between the analogous motifs in the urease-pyrimidinase superfamily (18). Thus, in spite of the resemblance between the conserved motifs, which may reflect similar aspects of the catalytic mechanisms, detailed sequence comparisons did not reveal a specific relationship between these two superfamilies of metal-dependent hydrolases. Multiple alignment-based database threading did not yield any significant hits for the PHP superfamily, but interestingly, the best hit, albeit with a relatively low score, was urease (2kau), which has a distorted TIM barrel fold and belongs to the urease-pyrimidinase superfamily (18). The relevance of this observation is hard to ascertain. It may indicate that the PHP superfamily has a distinct version of the TIM barrel fold, but a novel fold clearly remains a possibility.

The evidence for the phosphoesterase activity of the PHP domain comes from the yeast histidinol phosphatase encoded by the his-2 gene (20,21). This enzyme is involved in the ninth step of histidine biosynthesis in yeast and cleaves the phosphomonoester bond in histidinol phosphate to release histidinol, which is then reduced to histidine by the his-4c gene product (20). Several mutations resulting in a histidine-requiring phenotype have been identified in the his-2 gene (21), and interestingly, two of them map to the conserved motifs described above. The his2-6 mutation affects motif I and is situated directly N-terminal of the conserved HXH signature, whereas the his2-390 mutation affects the conserved histidine in motif 3 (Fig. 1). The mutation data are compatible with the conclusion that these conserved motifs are required for the enzyme function.

Using the scores produced by the gapped BLAST program as a similarity measure, the PHP superfamily was clustered into five distinct protein families, which may include phosphoesterases with distinct functions (Figs 1 and 2): 1) N-terminal domains of the bacterial DNA polIII α subunits, which consist of ~210 amino acids; 2) a group of stand-alone proteins from bacteria and archaea, which vary in size from 154 (so far the minimal size of the PHP domain) to 274 residues and show greater similarity to the polIII N-terminal domains than other members of the super-

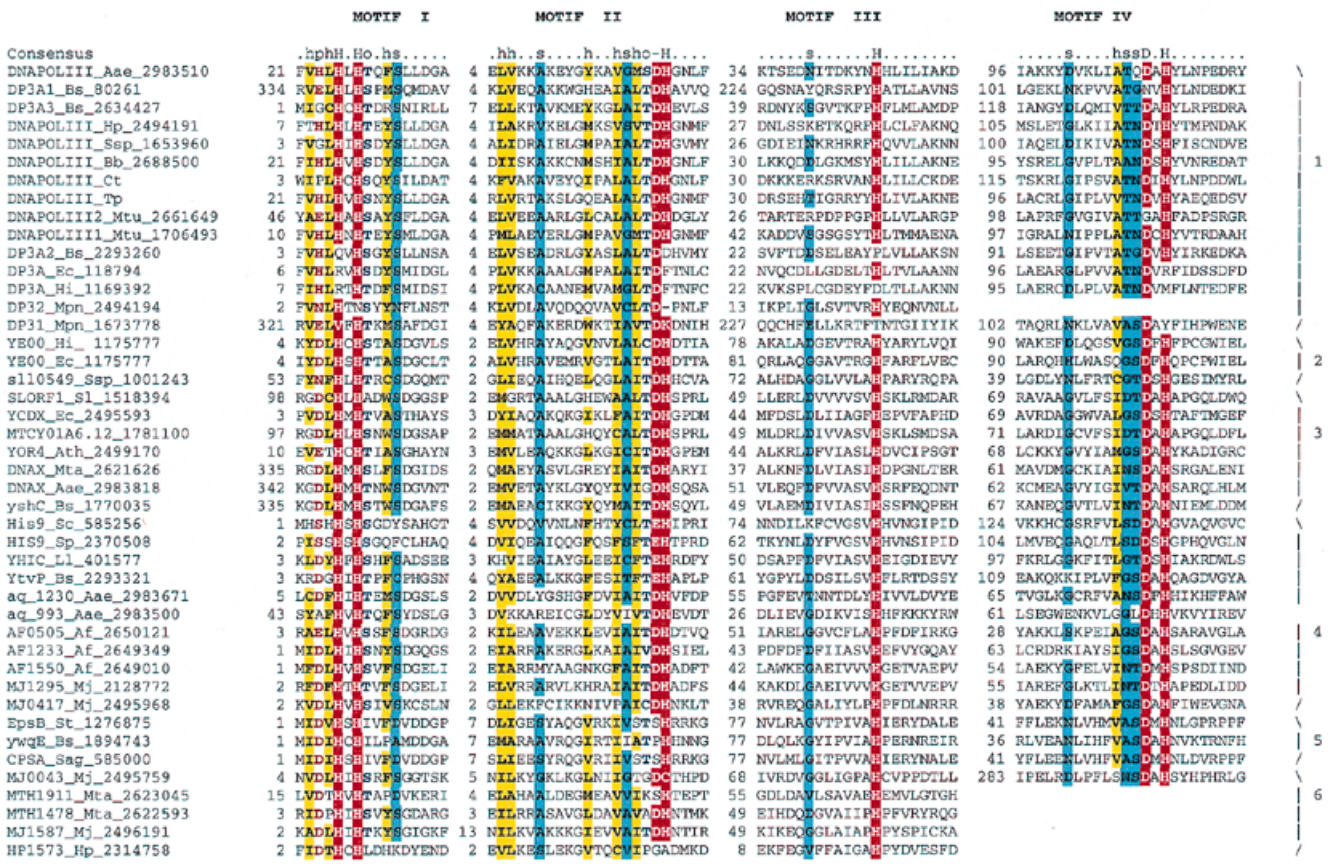

Figure 1. The four conserved motifs in the PHP domain. The alignment was constructed using the MACAW program and adjusted on the basis of the PSI-BLAST search results. The numbers indicate the distances from the protein N-termini to the first aligned block and the distances between the blocks. The sequences are grouped by similarity-based clusters identified using the CLUS program: 1) N-terminal domains of the bacterial DNA polIII α subunits; 2) group of stand-alone proteins from bacteria and archaea that show greater similarity to the polIII N-terminal domains than other members of the superfamily; 3) C-terminal domains of family X DNA polymerases from *B.subtilis* and *M.thermoautotrophicum* and highly similar stand-alone proteins from bacteria and archaea; 4) histidinol phosphatases from two yeast species and their apparent orthologs from *L.lactis* and *B.subtilis*; 5) a group of proteins encoded in operons involved in capsular biosynthesis in *Streptococcus*, *Staphylococcus* and *B.subtilis*. Group 6 includes all sequences lacking a detectable motif IV. The shading is based on a 85% consensus. The consensus is shown above the alignment; h indicates hydrophobic residues (A,C,F,I,L,M,V,W,Y; yellow background), s indicates small residues (A,C,S,T,D,N,V,G,P; blue background), p indicates polar residues (D,E,H,K,N,Q,R,S,T; brown coloring), o indicates hydroxy residues (S,T; green coloring), and '–' indicates negatively charged residues (D,E; purple coloring). The putative active site residues or those involved in metal chelation as described in the text are shown by inverse red shading. The yeast histidinol phosphatase encoded by the his-2 gene is designated HIS9 after the SWISS-PROT database since this is the enzyme catalyzing the ninth step of histidine biosynthesis. The species abbreviations are: Aae, *A.aeolicus*; Af, *Archaeoglobus fulgidus*; Ath, *Anaerocellum thermophilum*; Bb, *Borrelia burgdorferi*; Bs, *B.subtilis*; Ct, *Chlamydia trachomatis*; Ec, *E.coli*; Hi, *H.influenzae*; Hp, *Helicobacter pylori*; Ll, *L.lactis*; Mj, *Methanococcus jannaschii*; Mpn, *Mycoplasma pneumoniae*; Mta, *M.thermoautotrophicum*; Mtu, *Mycobacterium tuberculosis*; Mpn, *Mycoplasma pneumoniae*; Pf, *Pyrococcus furiosus*; Ph, *Pyrococcus horikoshii*; Pv, *Phaseolus vulgaris*; Sag, *Streptococcus agalactiae*; St, *Streptococcus thermophilus*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Ss, *Synechocystis* sp; Tp, *Treponema pallidum*.

family; 3) C-terminal domains of family X DNA polymerases from *B.subtilis*, *A.aeolicus* and *M.thermoautotrophicum* and highly similar stand-alone proteins from bacteria and archaea; 4) histidinol phosphatases from two yeast species and their apparent orthologs from *Lactococcus lactis* and *B.subtilis* (for the *L.lactis* protein, this functional prediction is compatible with its location in the vicinity of and probably within the same operon with other histidine biosynthesis genes) (22); 5) a group of proteins encoded in operons involved in capsular biosynthesis in *Streptococcus*, *Staphylococcus* and *B.subtilis* (23).

The presence of the PHP domains in two distinct polymerase families, especially its ubiquity in the α-subunits of PolIII, is of particular interest and may have important functional implications. The polymerase reaction involves the transfer of a nucleotide monophosphate from a deoxynucleoside triphosphate to a primer, accompanied by the release of inorganic pyrophosphate. The equilibrium of this reaction is shifted in the direction of polymerization if the pyrophosphate is removed as it is produced by the polymerase reaction. Pyrophosphate hydrolysis coupled with polymerization is thought to be catalyzed by a phosphatase that has not yet been definitively identified in any system; inorganic pyrophosphatase, which is one of the most abundant phosphatases in all cells, has been tentatively implicated (1). We hypothesize that the PHP domain fused with the domain responsible for nucleotide polymerization may carry out this phosphatase function in bacterial polIII. PolIII α-subunits from a wide phylogenetic spectrum of bacteria, including *Aquifex*, *Deinococcus*, *Chlamydia*, *Spirochaetaceae*, *Helicobacter*, *Synechocystis*, and at least one copy in the Gram-positive bacteria, contain all the conserved residues of the PHP domain that are predicted to participate in the phosphatase

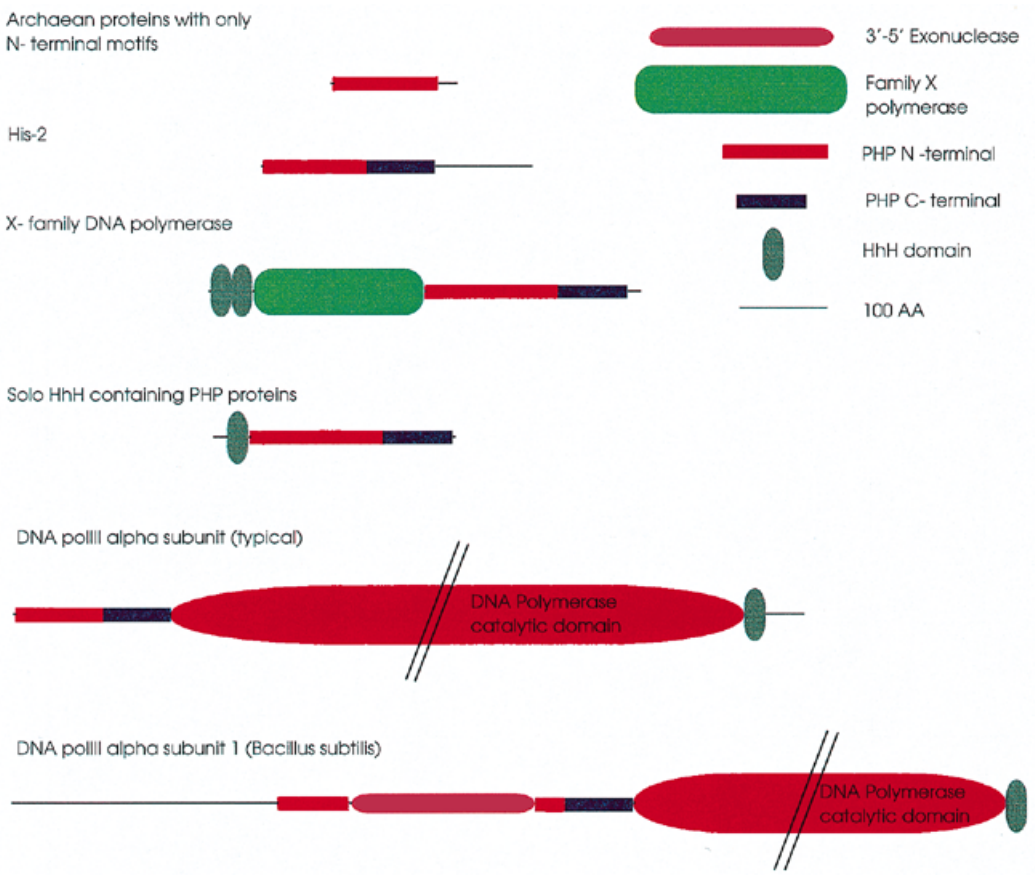

Figure 2. Distinct domain architectures of proteins containing the PHP domain. The figure is roughly to scale, except for the large DNA polymerase III α subunits. PHP N-terminal and PHP C-terminal refers to the N-terminal portion of the PHP domain (motifs I–III) and the C-terminal part (motif IV), respectively. HhH indicates a helix–hairpin–helix DNA-binding domain (35).

reaction (Fig. 1). However, in the polymerase from *Proteobacteria*, such as *Rickettsia*, *Escherichia coli* and *Haemophilus influenzae*, some of these residues are replaced, and in the case of *Mycoplasma*, some of the conserved motifs are additionally disrupted by deletions (Fig. 1). In Gram-positive bacteria, there are multiple copies of the polIII α-subunit. In *Mycobacteria*, both copies have intact proposed catalytic sites, but in one of the three *Bacillus* polymerases, they appear inactivated, and in the *Mycoplasma*, both copies are predicted to be inactive. Even though the PHP domain in *E.coli* polIII is predicted not to possess phosphatase activity, deletion mutagenesis experiments have shown that the N-terminal 60 amino acids are strictly required for the polymerase activity (17). It seems likely that the apparently inactive copies of the PHP domain perform an allosteric regulatory function, perhaps including pyrophosphate binding. Notably, *Proteobacteria* contain stand-alone versions of the PHP domain (e.g. *E.coli* YE00) that are predicted to possess the phosphatase activity and might supply this function to polIII *in trans*.

The PHP domain in the X-family polymerases may have the same function as proposed for the PolIII α subunit. However, this may not be the only explanation for their presence. Given the possibility that these DNA polymerases could be involved in repair, it cannot be ruled that in this case, the phosphoesterase domains function as nucleases. In the case of the PolIII α subunits, a functional nuclease is unlikely given the lack of evidence for any nuclease activity except for that of the well characterized 3′–5′ exonuclease domain (Fig. 2), in spite of the detailed experimental studies on this protein (1,16).

The small subunits of the archaeal DNA polymerases and eukaryotic DNA polymerases α and δ are members of the calcineurin-like phosphoesterase family

The small subunits of the eukaryotic DNA polymerases α, δ and ε showed moderate but statistically significant sequence similarity to their apparent orthologs from the four completely sequenced archaeal genomes (in a typical search, an e-value below 10^{-3} was observed in the second or the third iteration, but some eukaryotic polymerase subunits showed a greater similarity to the archaeal homologs; for example, the human polymerase δ subunit produced an e-value below 10^{-4} in the first pass of the search). Inspection of the alignment of these archaeal proteins revealed four conserved motifs, which define the vast superfamily of calcineurin-like phosphoesterases (19,24,25; Fig. 3). These motifs contain conserved histidine and aspartate residues involved in metal coordination and catalysis (24). These catalytic motifs, however, are disrupted in the eukaryotic polymerase subunits, in spite of the co-linear alignment throughout the entire eukaryotic and archaeal sequences (Fig. 3). These disruptions notwithstanding, when iterative PSI-BLAST searches were performed with the eukaryotic polymerase subunit sequences as queries, members of the phosphoesterase superfamily were retrieved from the NR database

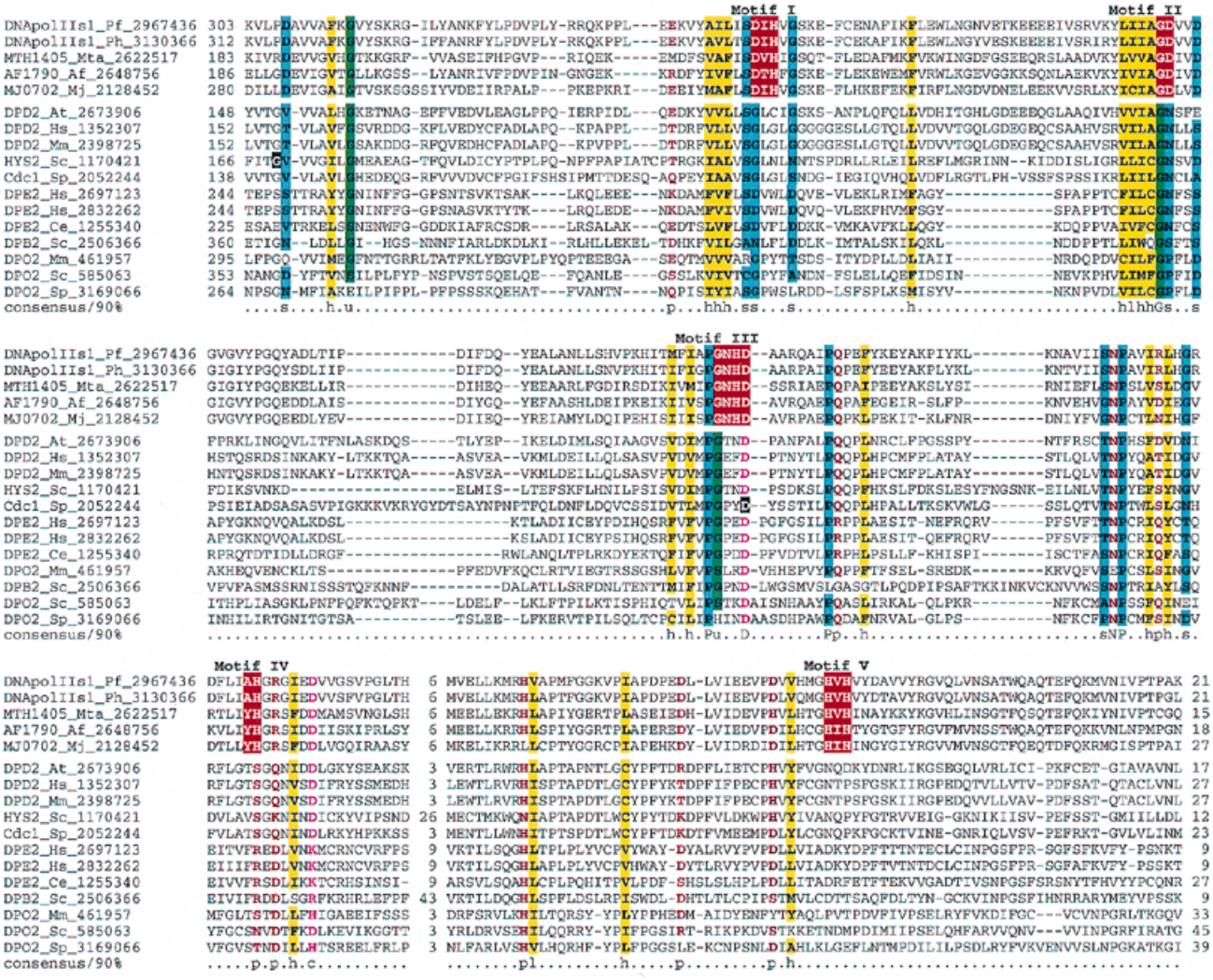

Figure 3. A complete alignment of the archaeal and eukaryotic DNA polymerase small subunits. The consensus was derived under the 90% rule. The designations and the convention used in shading are as in Figure 1; additionally, c indicates charged residues (D,E,K,R,H; magenta coloring) and u indicates 'tiny' residues (G,A,S; green background). The conserved motifs typical of the calcineurin-like phosphatase superfamily are indicated, and the five predicted catalytic and/or metal-chelating motifs in the archaeal polymerases (top five lines) and the residues mutated in the two Hys2 alleles are shown by inverse type. The protein name abbreviations: DPD2, DNA polymerase δ small subunit; DPO2, DNA polymerase α small subunit. The archaeal species abbreviations are as in Figure 1. Eukaryotic species abbreviations: At, *Arabidopsis thaliana*; Ce, *Caenorhabiditis elegans*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Sc, *S.cerevisiae*; Sp, *S.pombe*.

at statistically significant levels [for example, in a search initiated with the sequence of mouse polymerase δ small subunit (26), an Icc family member from *M.autotrophicum*, MTH1722 (Fig. 4), was retrieved with an e-value below 10^{-3} in the fourth iteration].

To evaluate the phosphoesterase connection statistically and to determine the details of the phylogenetic distribution, we constructed a profile based on the multiple alignment of the known members of the calcineurin-like superfamily and carried out a systematic search of the NR database and available complete genomes. This procedure resulted in the retrieval of the entire superfamily, including the small subunits of the eukaryotic and archaeal polymerases, at a statistically significant level (at least $e = 10^{-4}$ at the third iteration). In order to assess the robustness of the observed relationships, multiple searches were performed as described above for the PHP superfamily. The three-dimensional structure of calcineurin and the purple acid phosphatase have been determined (24,25), and the observed pattern of residue conservation in the polymerase subunits as well as the results of alignment-based secondary structure prediction are fully compatible with the presence of a calcineurin-like phosphoesterase core domain (Fig. 4).

The calcineurin-like superfamily consists of enzymes with diverse functions, including protein phosphoserine phosphatases, nucleotidases, sphingomyelin phosphodiesterases and 2′–3′ cAMP phosphodiesterases, as well as nucleases such as bacterial SbcD or yeast MRE11 (19,27). The superfamily is defined by five conserved blocks, which center around the metal-chelating residues; the four most conserved blocks are shown in Figure 4. We detected all the members of this superfamily from completely sequenced genomes and classified them into groups on the basis of sequence similarity; the polymerase subunits belong to a family typified by the bacterial Icc-like proteins whose functions are not known (Fig. 4). Given the absence of evidence for nuclease activity associated with the accessory subunits of archaeal or

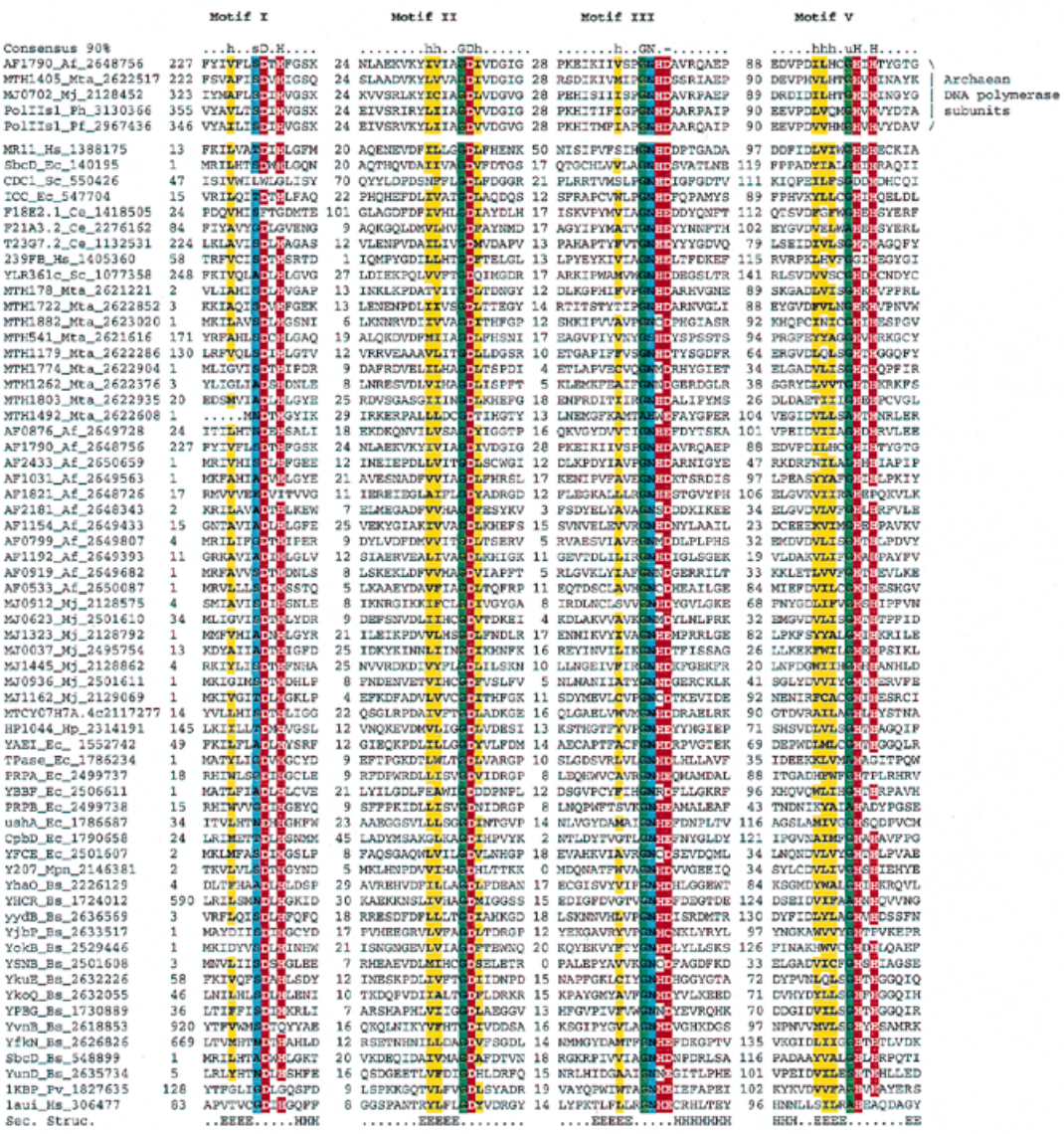

Figure 4. The four most conserved motifs in the calcineurin-like phosphoesterase superfamily. The rules for consensus derivation and shading are the same as in Figures 1 and 3. A fifth, short conserved motif was omitted for compactness. All the representatives of this superfamily from the three completely sequenced archaeal genomes as well as *E.coli* and *B.subtilis* are shown in addition to the archaeal DNA polymerase subunits (the top three lines), in order to indicate the diversity and the comparable representation of this superfamily in archaea and bacteria. The bottom two lines are the sequences of purple acid phosphatase from *P.vulgaris* (Pv) and human calcineurin, for which X-ray structures have been determined. The consensus secondary structure derived by comparison of these two structures is shown underneath the alignment; E indicates extended conformation (β-strand) and H indicates α-helix.

eukaryotic DNA polymerases (1,5), it seems likely that similarly to the PHP domain in bacterial polIII, the archaeal polymerase subunits possess phosphatase activity and may be required for pyrophosphate hydrolysis as discussed above. Recent studies on archaeal DNA polymerases have shown that the orthologs of the eukaryotic small subunits are in fact the small subunits of a novel DNA polymerase (designated DNA polymerase II) whose large, catalytic subunit is unrelated to any other known polymerases (28,29). Notably, it has been demonstrated that the archaeal DNA polymerase II possesses a 3′–5′ exonuclease activity that requires a complex of the two subunits. These striking new findings call for caution in the interpretation of the exact nature of the predicted phosphoesterase activity, and make experimental analysis of these proteins particularly interesting.

The eukaryotic DNA polymerase subunits with disrupted catalytic motifs have clearly lost the enzymatic activity. In mammalian DNA polymerase δ, the small subunit is required for the interaction with the clamp protein PCNA (30). The high conservation of PCNA in archaea (31) suggests that this interaction is an ancestral function in archaea and eukaryotes, and accordingly, the archaeal DNA polymerase small subunits may possess dual function, namely pyrophosphate hydrolysis, resulting in increased polymerization rate and PCNA binding. In eukaryotes, the phosphatase function may have been lost due to displacement by alternative phosphatases, but selection for the interaction with the clamp subunit could have resulted in the conservation of a structurally intact protein without enzymatic activity. Recently, the yeast polymerase δ small subunit, which is encoded by the

Hys2 gene, has been characterized (32). The mutation in the Hys2-2 allele affects the conserved aspartate in motif II of the phosphatase superfamily (Fig. 3). This suggests that despite the disruption of most of the predicted metal-chelating sites in the eukaryotic polymerase subunits, coordination of a divalent cation by this aspartate together with the conserved histidine in motif V could be important to maintain the structure of these proteins. Similar functions may be proposed for the less thoroughly characterized small subunits of DNA polymerases α and ε that are paralogs of the polymerase δ small subunit (33 and the present work) and like the latter, appear to be an inactivated phosphatases (Fig. 3).

Functional and evolutionary implications of the association of phosphoesterase domains with DNA polymerases

The analysis described here resulted in the identification of two unrelated phosphoesterase domains in four distinct types of DNA-dependent DNA polymerases, namely: (i) the N-terminal PHP domain in bacterial polIII α subunits; (ii) the C-terminal PHP domain in bacterial and archaeal X family DNA polymerases; (iii) an apparently active calcineurin-like phosphatase superfamily domain in the small subunit of archaeal DNA polymerase II; and (iv) an inactivated calcineurin-like phosphatase superfamily domain in the small subunits of eukaryotic DNA polymerases α, δ and ε. A striking parallel between the PHP domain and the calcineurin-like phosphatase domain is the existence, in addition to the forms containing all predicted catalytic residues, of apparently inactivated versions. The homologous relationship between the domains predicted to possess phosphatase activity and the inactivated ones is strongly supported statistically. Given the multiple motifs that are required for catalytic activity of phosphoesterases but are disrupted in the apparently inactive forms (e.g. Fig. 3), it is clear that in evolutionary terms, the active enzyme is the ancestral state. It appears most likely that in both instances, the polymerase–phosphatase association has evolved through the recruitment of an active phosphatase. For the PHP domain, this scenario is further supported by the presence of a domain predicted to possess phosphoesterase activity in the polIII α subunits from all major bacterial lineages, including the ancient *Aquifex* branching, whereas the apparently inactive forms are found in only two, namely Proteobacteria and Gram-positive bacteria (Fig. 2). In the case of the eukaryotic polymerase subunits, it appears that there had been an ancestral inactivation of the calcineurin-like phosphoesterase domain, followed by a series of duplications resulting in the extant subunits of the α, δ and ε polymerases.

Fusion of phosphoesterase domains with polymerases is not limited to the cases discussed here. We had previously described a fusion of the polyA polymerase domain (polymerase X superfamily) to a newly defined phosphoesterase domain called the DHH domain (34). In addition, we also noticed the fusion of another novel phosphoesterase domain to the tRNA-CCA adding enzyme, also of the polymerase X family (L.A. and E.V.K., unpublished observations). These multiple and obviously independent fusions suggest some selective advantage(s) conferred by the phosphoesterase–polymerase association. Pyrophosphate hydrolysis resulting in a shift of the reaction equilibrium towards nucleotide polymerization is a plausible unifying explanation that deserves experimental testing, though alternative interpretations also should be considered. Specifically, the possibility exists that some of the phosphoesterase domains actually possess nuclease activity. Furthermore, the independent disruption of the motifs implicated in catalysis in two classes of polymerase-associated phosphoesterase domains, whose function, however, remains essential for the polymerase activity, may even suggest that their primary role may be non-enzymatic. Such roles may include regulation of polymerase activity, perhaps through pyrophosphate binding and/or interaction with other subunit of the holoenzyme, like the clamp subunit in archaeal and eukaryotic polymerases. Experimental elucidation of the roles of active and inactivated phosphoesterase domains in polymerases will hopefully reveal fundamental but hitherto unnoticed aspects of DNA replication in all known systems, and clarify the relationship between the predicted enzymatic and regulatory functions of these domains.

REFERENCES

1 Kornberg,A. and Baker,T. (1992) *DNA replication*. 2nd edition, WH Freeman, San Francisco.
2 Friedberg,E.C., Walker,G.C. and Siede,W. (1995) *DNA Repair and Mutagenesis*. ASM Press, Washington, DC.
3 Braithwaite,D.K. and Ito,J. (1993) *Nucleic Acids Res.*, **21**, 787–802.
4 Mushegian,A.R. and Koonin,E.V. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 10268–10273.
5 Hindges,R. and Hubscher,U. (1997) *Biol. Chem.*, **378**, 345–362.
6 Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.
7 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
8 Schuler,G.D., Altschul,S.F. and Lipman,D.J. (1991) *Proteins*, **9**, 180–190.
9 Neuwald,A.F., Liu,J.S. and Lawrence,C.E. (1995) *Protein Sci.*, **4**, 1618–1632.
10 Tatusov,R.L., Altschul,S.F. and Koonin,E.V. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 12091–12095.
11 Koonin,E.V., Tatusov,R.L. and Rudd,K.E. (1996) *Methods Enzymol.*, **266**, 295–322.
12 Rost,B. and Sander,C. (1994) *Proteins*, **19**, 55–72.
13 Rost,B., Schneider,R. and Sander,C. (1997) *J. Mol. Biol.*, **270**, 471–480.
14 Wootton,J.C. and Federhen,S. (1996) *Methods Enzymol.*, **266**, 554–573.
15 Holm,L. and Sander,C. (1995) *Trends Biochem. Sci.*, **20**, 345–347.
16 Yue,D., Maizels,N. and Weiner,A.M (1996) *RNA*, **2**, 895–908.
17 Kim,D.R., Pritchard,A.E. and McHenry,C.S. (1997) *J. Bacteriol.*, **179**, 6721–6728.
18 Holm,L. and Sander,C. (1997) *Proteins*, **28**, 72–82.
19 Koonin,E.V. (1994) *Protein Sci.*, **3**, 356–358.
20 Jones,E.W. and Fink,G.R. (1982) In Strathern,J.N., Jones,E.W. and Broach,J.R. (eds), *The Molecular Biology of the Yeast Saccharomyces*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Vol. 2, pp. 264–299.
21 Malone,R.E., Kim,S., Bullard,S.A., Lundquist,S., Hutchings-Crow,L., Cramton,S., Lutfiyya,L. and Lee,J. (1994) *Genetics*, **137**, 5–18.
22 Delorme,C., Ehrlich,S.D. and Renault,P. (1992) *J. Bacteriol.*, **174**, 6571–6579.
23 Roberts,I.S. (1996) *Annu. Rev. Microbiol.*, **50**, 285–315.
24 Goldberg,J., Huang,H.B., Kwon,Y.G., Greengard,P., Nairn,A.C. and Kuriyan,J. (1995) *Nature*, **376**, 745–753.
25 Strater,N., Klabunde,T., Tucker,P., Witzel,H. and Krebs,B. (1995) *Science*, **268**, 1489–1492.
26 Zuo,S., Gibbs,E., Kelman,Z., Wang,T.S., O'Donnell,M., MacNeill,S.A. and Hurwitz,J. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 11244–11249.
27 Sharples,G.J. and Leach,D.R. (1995) *Mol. Microbiol.*, **17**, 1215–1217.
28 Uemori,T., Sato,Y., Kato,I., Doi,H. and Ishino,Y. (1997) *Genes Cells*, **2**, 499–512.
29 Ishino,Y., Komori,K., Cann,I.K. and Koga,Y. (1998) *J. Bacteriol.*, **180**, 2232–2236.
30 Zhou,J.Q., He,H., Tan,C.K., Downey,K.M. and So,A.G. (1997) *Nucleic Acids Res.*, **25**, 1094–1099.
31 Brown,J.R. and Doolittle,W.F. (1997) *Microbiol. Mol. Biol. Rev.*, **61**, 456–502.
32 Hashimoto,K., Nakashima,N., Ohara,T., Maki,S. and Sugino,A. (1998) *Nucleic Acids Res.*, **26**, 477–485.
33 Li,Y., Asahara,H., Patel,V.S., Zhou,S. and Linn,S. (1997) *J. Biol. Chem.*, **272**, 32337–32344.
34 Aravind,L. and Koonin,E.V. (1998) *Trends Biochem. Sci.*, **23**, 17–19.
35 Doherty,A.J., Serpell,L.C. and Ponting,C.P. (1996) *Nucleic Acids Res.*, **24**, 2488–2497.