Invariant Asp-1122 and Asp-1124 Are Essential Residues for Polymerization Catalysis of Family D DNA Polymerase from *Pyrococcus horikoshii*

Received for publication, December 28, 2000, and in revised form, April 13, 2001
Published, JBC Papers in Press, April 23, 2001, DOI 10.1074/jbc.M011762200

Yulong Shen‡, Krishnasasya Musti‡, Madoka Hiramoto‡, Hisasi Kikuchi‡, Yutaka Kawarabayashi‡, and Ikuo Matsui‡‡

From the ‡National Institute of Bioscience and Human Technology, Tsukuba, Ibaraki 305, Japan and the §National Institute of Technology and Evaluation, Ministry of International Trade and Industry, Nishihara, Shibiyaku, Tokyo 150, Japan

Family D DNA polymerase has recently been found in the Euryarchaeota subdomain of Archaea. Its genes are adjacent to several other genes related to DNA replication, repair, and recombination in the genome, suggesting that this enzyme may be the major DNA replicase in Euryarchaeota. Although it possesses strong polymerization and proofreading activities, the motifs common to other DNA polymerase families are absent in its sequences. Here we report the mapping of the catalytic residues in a family D DNA polymerase from *Pyrococcus horikoshii*. Site-directed alanine mutants for 28 conserved aspartic acid or glutamic acid residues were screened for polymerization and 3′–5′ exonuclease activities. We identified the invariant aspartates Asp-1122 and Asp-1124 within the most conserved motif as the catalytic residues involved in DNA polymerization. Alanine mutation at either site caused a loss of polymerization activity, whereas the conserved mutants, D1122E, D1124N, and D1124E, had slightly reduced polymerization activity. We also found that the 3′–5′ exonuclease activity remains in D1122A and D1124A, indicating that the catalytic residues of DNA polymerization are different from those of the 3′–5′ exonuclease activity. Furthermore we determined the molecular mass of the recombinant enzyme by gel filtration and proposed a heterotetrameric structure for this enzyme.

DNA polymerases have been classified into four major families, A, B, C, and D, represented by *Escherichia coli* DNA polymerase I, II, III α subunit, and *Pyrococcus furiosus* polymerase II, respectively, based on alignment of amino acid sequences (1–3). Family D DNA polymerase (PolD) † is a new family recently found in the Euryarchaeota subdomain of Archaea, the third domain of life (4–7). So far at least eight PolDs have been reported, and their sequences have been deposited in the genetic data bases. PolDs are composed of two subunits: a small subunit (DP1), which shows low but significant homology to eukaryotes DNA polymerase δ (7), and a large subunit (DP2), which is believed to be the catalytic subunit because of its weak polymerization activity (5). The interaction between the two subunits is essential for full enzyme activities. Strong polymerization and 3′–5′ exonuclease activities are obtainable only when the two subunits were mixed or co-expressed together (5). Amino acid sequences of the large subunit have no homology to other DNA polymerases and do not contain common motifs of other DNA polymerase families despite the fact that members of PolD from the Euryarchaeota species are highly conserved. PolD has been mainly characterized in *P. furiosus*; nevertheless, many aspects of the enzyme properties, such as the catalytic residues for both polymerization and 3′–5′ exonuclease activities, are still unknown.

It is of particular interest that the genes coding the two subunits of PolD in *Pyrococcus* (*P. furiosus, P. horikoshii, P. abyssi, and P. woesei*) exist in an operon containing three or four other ORFs that are important in DNA replication, DNA repair, and recombination (6, 8). One upstream ORF (PH0124 in *P. horikoshii*) is a homologue to Orc1 (subunit 1 of the origin recognition complex) of yeast, and a downstream ORF (PH0119) is a homologue to yeast RadB, a protein involved in recombination. Moreover, this operon is located adjacent to a recently identified replication origin in the *Pyrococcus* genomes (8). The genes of replication factor C (ORF PH0112 and PH0113 in *P. horikoshii*) and putative DNA helicase (PH0109) are located only 6–10 kilobase pairs away from the genes of PolD. This clustering of genetically essential genes conserved in *Pyrococcus* indicates that PolD is likely the main replicase in DNA replication in Euryarchaeota. Because the enzymes involved in DNA replication in Archaea are similar to the replicative enzymes in eukaryotes (9), further characterization of the proteins related to DNA replication, including PolD, will help reveal the DNA replication mechanism in Archaea, as well as that in eukaryotes, which has become complicated during evolution and is difficult to study. Fortunately, the thermostability of the replication proteins and clustering of the replication genes facilitate their analysis. In the present study, we successfully cloned, expressed, and purified the PolD from *P. horikoshii* (PolDPho). By site-directed mutagenesis, we identified amino acid residues Asp-1122 and Asp-1124 as the essential residues responsible for DNA polymerizing activity. Using gel filtration analysis, we determined the molecular mass of the recombinant PolD from *P. horikoshii* and proposed that PolD from *P. horikoshii* probably forms a heterotetrameric structure in solution.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ To whom correspondence should be addressed: National Institute of Bioscience and Human Technology, Tsukuba, Ibaraki 305, Japan. Tel.: 81-298-616142; Fax: 81-298-616151; E-mail: imatsui@nibh.go.jp.
† The abbreviations used are: PolD, family D DNA polymerase; DP1, small subunit of the family D DNA polymerase; DP2, large subunit of the family D DNA polymerase; PolDPho, family D DNA polymerase from *P. horikoshii*; DP1Pho, small subunit of the family D DNA polymerase from *P. horikoshii*; DP2Pho, large subunit of the family D DNA polymerase from *P. horikoshii*; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction.
isopropylthio-

amplified fragment was digested with

amplified using PCR with primers S1 and S2 (Table I; Fig. 1). The
gene for the small subunit of PolD from

P. horikoshii

chased from Roche Molecular Biochemicals.

Promega, and protease inhibitor mixture tablets (EDTA-free) were pur-
purchased from Takara Shuzo and was used according to the manufac-
Takara Shuzo (Shiga, Japan) and Promega. The DNA ligation kit was

land Biolabs (Beverly, MA). Restriction enzymes were purchased from

from Stratagene. Vent DNA polymerase was purchased from New Eng-
strain BL21-CondonPlus (DE3)-RIL competent cells were purchased

MRF

cells were purchased from Stratagene (La Jolla, CA). The pGE-
P. coli

ampicillin (100 mg/liter) at 37 °C. When

A

transformed into host

elements of the genes were verified using the LIC-4200L(S)-2 DNA

sequences of the genes were verified using the LIC-4200L(S)-2 DNA

final co-expression vector was named pET15b/SL (Fig. 1). The se-
pGEMEX-1 were then subcloned sequentially into pET15b-SM. The

I; Fig. 1). The DP2Pho gene fragments together with a sequence from

Sac

ribosomal binding site using PCR, a primer (pGEM-rbs) containing a

Table I). The resulting plasmid was named pET15b-SM. To include the

site in pET15b-S was modified with PCR using primers S1 and

L2. The lower part of the DP2Pho gene (DP2L, from 2348 to 4305

base pairs) was amplified using primers L3 and L4. After digestion with

the restriction enzymes, DP2U and DP2L were inserted into the

gEMEX-1 vector sequentially to generate a vector pGEMEX-L(

pGEMEX-1 vector sequentially to generate a vector pGEMEX-L(

NdeI and BamHI and inserted into an expression vector pET15b (Novagen, Madison, WI) to produce a

vector named pET15b-S. The DP2Pho gene with the intein was split

into two parts using PCR (Table I; Fig. 1). The upper part of the DP2Pho
gene (DP2U, from 1 to 2347 base pairs) was amplified with primers L1

and L2. The lower part of the DP2Pho gene (DP2L, from 2348 to 4305

base pairs) was amplified using primers L3 and L4. After digestion with

the restriction enzymes, DP2U and DP2L were inserted into the

gEMEX-1 vector sequentially to generate a vector pGEMEX-L(

L1 5

9

9

-GCTTGTCGACGCCATAAACTTTGACATTATCCATTGCGCGCTTAAGCAAC-3

9

9

-TTTATGGCGTCGACAAGCTGAAGG-3

L3 5

9

9

-GTTGTCGACGCCATAAACTTTGACATTATCCATTGCGCGCTTAAGCAAC-3

9

9

-CTCGACTTTAGCATATGGCGCTGATGGAGC-3
Polymerization of Family D DNA Polymerase

checked by SDS-PAGE performed on a 10–15% gradient gel using the Phast system (Amersham Pharmacia Biotech). Protein bands were visualized by staining with Coomassie Brilliant Blue R-250. The supernatant was dialyzed against buffer A (50 mM Tris-HCl, pH 8.0) and then against buffer B (50 mM Tris-HCl, pH 6.5). The dialysate was loaded onto a hydrophobic column (Amersham Pharmacia Biotech) equilibrated with buffer B using the fast protein liquid chromatography system (Amersham Pharmacia Biotech). The column was developed with a linear gradient of 0–1000 mM. Fractions at around 450 mM NaCl, which contained the PolDPho complex, were dialyzed against buffer A to remove the salt. Proteins from pET15b/S and mutant plasmids were further purified by loading them onto a nickel column (Novagen). After washing with a buffer containing 20 mM NaCl, the enzymes were eluted with the elution buffer containing 150 mM imidazole. The protein content was determined with the Bio-Rad protein assay dye reagent using bovine serum albumin as the standard protein.

Native Molecular Mass Determination—The native molecular mass was estimated using fast protein liquid chromatography gel filtration on a Hichain Superdex 200-μg (26/60) column (Amersham Pharmacia Biotech) equilibrated with 50 mM Tris-HCl buffer (pH 7.0) and 200 mM NaCl. The fraction size was 5.0 ml, and the flow rate was 2.0 ml/min. The standards used were 669-kDa thyroglobulin, 440-kDa ferritin, 232-kDa catalase, 158-kDa aldolase, and 67-kDa bovine serum albumin.

Assay of DNA Polymerizing Activity—The presence of dNTP in the reaction mixture is essential to synthesize DNA products by PolD. In the absence of dNTP, the substrate DNA is digested by the exonuclease activity, and the reaction mixture contains both the exonuclease activity is measurable. The DNA polymerizing activity was assayed by measuring the incorporation of [α-32P]ATP into the trichloroacetic acid-insoluble material. The standard reaction mixture (in 50 μl) contained 20 mM Tris-HCl (pH 8.8, 25 °C), 10 mM MgSO4, 10 mM KCl, 10 mM (NH4)2SO4, 0.1% Triton X-100, 200 μM heat-denatured salmon testes DNA, 0.25 mM dNTP mixture, 0.5 μCi (9.25 kBq) of [α-32P]ATP, 50 μg of labeled or purified enzymes. The reaction was performed at 70 °C for 30 min, and then the radioactivity was measured by scintillation counting. At least two repeats were performed independently for each sample.

Assay of Primer Extension Activity—The primer extension ability was assessed using a 100-mer oligonucleotide whose sequences were taken from M13mp19 single-stranded DNA (5′-CTCTTACTGAATCGACCTCGTCAAGGTTCGGAATCCCGCTTACATCAATCCGTTTCTCCAAAGCTTCTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT-3′, complementary to positions 6337–6304 of M13mp18) as a substrate. The 34-mer end-labeled 3′-phosphate was incubated with the bacteriophage T4 polynucleotide kinase and [γ-32P]ATP to make the substrate, 10 mM Tris-HCl buffer (pH 8.0) and 500 ng of heat-treated supernatant or 400 μl of 50 mM Tris-HCl buffer containing 200 μg/ml heat-denatured E. coli strain XL1 Blue (MRF−) was added, and the tube was mixed and boiled in boiling water. After 10 min, 10 μl of SSC buffer was added, and the mixture was heated at 60 °C for 10 min. The gapped DNA was recovered by ethanol precipitation and confirmed by gel electrophoresis. The gap was filled by PolDPho with polymerization for 30 min at 70 °C. The reaction was stopped by adding 1.5 mM EDTA, and the DNA was recovered by ethanol precipitation and dissolved in TE. The filling of the gap was confirmed by electrophoresis using agarose gel. E. coli strain XL1 Blue (MRF−) was transformed with the purified DNA, and strain MV1184 was used as lawn cells for transfection. White plaques were counted, and the fidelity was calculated as a percentage of the number of white plaques against the total plaques formed.

RESULTS

Cloning and Expression of the Small and Large Subunit Genes of PolDPho—According to the genomic sequences of P. horikoshii OT3 (13), the two ORFs PH0123 and PH0121 code for the small subunit DP1Pho (622 amino acids, 70 kDa) and large subunit DP2Pho (1434 amino acids, 163 kDa) of PolD, respectively. As in P. furiosus and P. abyssi, the two genes are arranged in tandem (with DP1Pho in the front) and are adjacent to ORFs that are important for DNA replication, recombination, and repair (6, 8). Alignment and sequence analysis of DP2Pho with other PolD large subunits from the Archaea species revealed that DP2Pho contains a 166-amino acid mini-intein (internal protein fragment) from Cys-955 to Gin-1120 (Fig. 2; Ref. 14). The large subunit without the intein is 1268 amino acids in length and has a predicted molecular mass of 144 kDa.

DP1Pho was cloned and expressed in E. coli strain BL21-CodonPlus (DE3)-RIL and its purification from soluble fractions were successful. The expression of DP2Pho using pGEMEX-L(−) and pGEMEX-L(+)) was also successful; however, the majority of the protein was in insoluble fractions. Purification of DP2Pho from soluble fractions proved difficult because of its insolubility and the degradation of the protein during purification.

To obtain a stable recombinant protein and active complex, we constructed the co-expression vector pET15b/S. In this vector, the DP1Pho and DP2Pho genes were connected in tandem into two ORFs. The two genes are transcribed under the control of a single transcription promoter and terminator (T7 promoter and T7 terminator). After transformation and induction using the E. coli strain BL21-CodonPlus (DE3)-RIL, stable proteins were expressed as assayed by SDS-PAGE and were purified to near homogeneity with an anion exchange column and nickel column. Several milligrams of purified protein could be produced from a liter of culture. Fig. 3 shows the SDS-PAGE and Western blotting profile of His-tagged PolDPho. The band positions of DP2Pho were as predicted at around 140 kDa, whereas DP1Pho migrated at around 90 kDa rather than the predicted 70 kDa. The band intensity of DP2Pho was about two times that of DP1Pho, as measured by density scanning of the protein bands on the SDS gel (data not shown), indicating a 1:1 molar ratio between the two subunits. Our success in purification of PolDPho with a histidine tag (His tag) at the N terminus of DP1Pho confirmed that DP1Pho and DP2Pho form a tight
The Thermoplasma acidophilum (accession number AE005116-6), and Halobacterium (Archaeoglobus fulgidus, accession number U67603-3), Methanococcus jannaschii MB.ther (Methanobacterium thermoautotrophicum, accession number D84670-3), P. horikoshii (P. horikoshii sp. NRC-1), and P. abyssi (P. abyssi, accession number AJ248283-120), P. furiosus P. abyssi (P. abyssi, accession number AP000001-125), P. horikoshii (P. horikoshii, accession number AE000913-9), Methanococcus jannaschii MB.ther (Methanobacterium thermoautotrophicum, accession number AE000984-2), Halobacterium (Archaeoglobus fulgidus, accession number U67603-3), and Thermus aquaticus (T. aquaticus, accession number AL445063-36). The identified residues for polymerization catalysis are indicated by arrowheads. The sequences were aligned using the CLUSTALW program.

The sequences are from eight Euryarchaeota species, indicate similar residues among species. The sequences were aligned using the CLUSTALW program.

FIG. 2. Alignment of partial amino acid sequences of DP2 in the region analyzed. The sequences are from eight Euryarchaeota species, P. horikoshii (P. horikoshii, accession number AP000001-125), P. abyssi (P. abyssi, accession number AJ248283-120), P. furiosus (P. furiosus, accession number D84670-3), Methanobacterium thermoautotrophicum (M.ther, accession number AE000913-9), Methanococcus jannaschii (M. jannaschii, accession number U67603-3), Archaeoglobus fulgidus (A. fulgidus, accession number AE000984-2), Halobacterium sp. NRC-1 (Hal.sp, accession number AE005116-6), and Thermus aquaticus (T. aquaticus, accession number AL445063-36). The asterisks indicate identical residues; the period and colon indicate similar residues among species. The identified residues for polymerization catalysis are indicated by arrowheads. The sequences were aligned using the CLUSTALW program.

Polymerization activities of the heat-treated proteins were examined by measuring the incorporation of \([\alpha-32P]dATP\) into salmon testes DNA. We found that proteins using the intein-containing vectors, pET15b/SL(+), and pET11a/SL(+), had only 10–20% of the DNA polymerization activity of proteins using vectors without the intein, pET15b/SL and pET11a/SL(−). This indicated that a small proportion of intein was self-spliced from the proteins. We also found that the addition of the His tag at the N-terminus of DP1Pho had no effect on the DNA polymerization activities. Therefore pET15b/SL was used as the co-expression vector for further expression and mutagenesis.

FIG. 3. SDS-PAGE (A) and Western blotting (B) of the purified PolDPho using the co-expression vector pET15b/SL constructed. A gel (10–15%) was used for both Western blotting with nickel-nitrilotriacetic acid alkaline phosphatase and staining with Coomassie Brilliant Blue R-250 after electrophoresis. M, the molecular mass standards: myosin, 205 kDa; β-galactosidase, 121 kDa; bovine serum albumin, 70 kDa; ovalbumin, 52 kDa; carbonic anhydrase, 34 kDa; soybean trypsin inhibitor, 29 kDa; lysozyme, 21 kDa; and aprotinin, 7 kDa.

Downloaded from http://www.jbc.org/ by guest on July 24, 2018

Measurement of Primer Extension and Exonuclease Activity of the Wild Type Enzyme—The primer extension ability and exonuclease activity of PolDPho were measured by gel filtration (Fig. 4). Purified protein (after anion exchange and nickel chromatography) was loaded onto a Superdex 200 pg column (Amersham Pharmacia Biotech). PolDPho was eluted at around 151 ml. By calculation with elution volumes of calibration markers (128 ml of thyroglobulin, 145 ml of ferritin, 171 ml of catalase, 176.3 ml of aldolase, and 193.5 ml of albumin), the molecular mass of PolDPho was determined to be 421 kDa. After gel filtration, each fraction was concentrated 20 times, and the concentrated samples were subjected to DNA polymerization activities. Therefore pET15b/SL was used as the co-expression vector for further expression and mutagenesis.

 Determination of the pH and Magnesium Ion Optima and Thermostability—The optimal pH and magnesium requirement for DNA synthesis were determined by measuring the incorporation of \([\alpha-32P]dATP\) into salmon testes DNA under different conditions or treatments. PolDPho showed the highest activity from pH 8.5 to 9.0, and the optimal magnesium concentration was 17.5 mM. PolDPho is highly thermostable; 50% activity remained after heating at 80 °C for 60 min. The CD profiles of PolDPho at the near UV region measured at 25 and 90 °C showed similar patterns, indicating no conformational alteration at high temperature (data not shown).

 Determination of Native Molecular Mass by Gel Filtration—The molecular mass of PolDPho was assessed by gel filtration (Fig. 4). Purified protein (after anion exchange and nickel chromatography) was loaded onto a Superdex 200 pg column (Amersham Pharmacia Biotech). PolDPho was eluted at around 151 ml. By calculation with elution volumes of calibration markers (128 ml of thyroglobulin, 145 ml of ferritin, 171 ml of catalase, 176.3 ml of aldolase, and 193.5 ml of albumin), the molecular mass of PolDPho was determined to be 421 kDa. After gel filtration, each fraction was concentrated 20 times, and the concentrated samples were subjected to DNA polymerization (by measuring the incorporation of \([\alpha-32P]dATP\) into salmon testes DNA, SDS-PAGE, and Western blotting analysis). The fraction activities varied coordinately with the UV absorption (Fig. 4) and protein band intensity (not shown).

To compare the effects of the intein and His tag on expression and activity, vectors containing the intein and vectors without the His tag were also constructed (named pET15b/SL(+), pET11a/SL(−), and pET11a/SL(+)) using a similar approach to that described above. Preliminary results showed that when using any of the three co-expression vectors, both DP1Pho and DP2Pho were also overexpressed in E. coli. The DNA polymerization activities of the heat-treated proteins were examined by measuring the incorporation of \([\alpha-32P]dATP\) into salmon testes DNA. We found that proteins using the intein-containing vectors, pET15b/SL(+), and pET11a/SL(+), had only 10–20% of the DNA polymerization activity of proteins using vectors without the intein, pET15b/SL and pET11a/SL(−). This indicated that a small proportion of intein was self-spliced from the proteins. We also found that the addition of the His tag at the N-terminus of DP1Pho had no effect on the DNA polymerization activities. Therefore pET15b/SL was used as the co-expression vector for further expression and mutagenesis.

FIG. 3. SDS-PAGE (A) and Western blotting (B) of the purified PolDPho using the co-expression vector pET15b/SL constructed. A gel (10–15%) was used for both Western blotting with nickel-nitrilotriacetic acid alkaline phosphatase and staining with Coomassie Brilliant Blue R-250 after electrophoresis. M, the molecular mass standards: myosin, 205 kDa; β-galactosidase, 121 kDa; bovine serum albumin, 70 kDa; ovalbumin, 52 kDa; carbonic anhydrase, 34 kDa; soybean trypsin inhibitor, 29 kDa; lysozyme, 21 kDa; and aprotinin, 7 kDa.
3′–5′ exonuclease activities of PolDPho are shown in Fig. 5. For the primer extension assay, 100 ng of enzyme was used in one reaction. The products (35–84-mer) were observed after reacting for 2 min (Fig. 5A). In the 3′–5′ exonuclease assay, more enzyme (400 ng) was used, and the reaction was performed for 30 min. The products were shorter than 34-mer (Fig. 5B).

**Measurement of Fidelity by Kunkel Method**—The fidelity of DNA polymerization by PolDPho was assessed using the Kunkel method. Among a total of 122,202 plaques counted, 109 (0.89%) were white. The fidelity of Taq DNA polymerase, which lacks 3′–5′ exonuclease activity, was assayed simultaneously as a control. The fidelity of Taq DNA polymerase was 1.65% (209/12,681) in our measurement, which was 1.9 times lower than PolDPho.

**Making Mutants to Map Catalytic Residues of Polymerization Activity**—Amino acid residues in the region from Glu-800 to Glu-1195 of DP2Pho were targeted for site-directed mutagenesis. This region is clustered with negatively charged residues conserved in DP2 (Fig. 2). In addition, it contains the most conserved and unknown motif, which is disrupted by the mini-intein in *P. horkoshii* (from Cys-955 to Gln-1120 of DP2Pho), *P. abyssi*, and *Halobacterium* sp. NRC-1 (Fig. 2). Two common DNA polymerase motifs were proposed in this region through visual inspection (Ref. 7; motif A, Lys-817–Phe-833, and motif C, Lys-884–Phe-897). The conserved aspartic acid through visual inspection (Ref. 7; motif A, Lys-817–Phe-833, and motif C, Lys-884–Phe-897). The conserved aspartic acid (Asp-818, Asp-824, Asp-827, Asp-853, Asp-872, Asp-891, Asp-892, Asp-911, Asp-1122, Asp-1124, Asp-1126, Asp-1135, Asp-1152, Asp-1162, Asp-1167, and Asp-1174), and glutamic acid (Glu-800, Glu-803, Glu-843, Glu-847, Glu-910, Glu-1125, Glu-1165, Glu-1169, Glu-1182, Glu-1185, and Glu-1195) residues in this region were changed to alanine (Table II). Mutant proteins were expressed and partially purified by heat filtration. Among the 28 mutants, only D1122A and D1124A showed loss of polymerization activity, whereas the other mutants did not show a dramatic change in DNA synthesis in the preliminary screening (Table II). Mutant proteins of D112A and D1124A were further purified by passing through anion exchange and nickel columns and the polymerization activities were examined by measuring the incorporation of [α-32P]dATP into salmon testes DNA, was based on the mean of two measurements with each fraction.

<Fig. 4. Determination of molecular mass of PolDPho by gel filtration. O and ● indicate UV absorption and polymerization activity of peak fractions of PolDPho by gel filtration, respectively. The eluting positions of the protein markers thyroglobulin (*T*θ, 669 kDa), ferritin (*Fe*, 440 kDa), catalase (*Cat*, 232 kDa), and aldolase (*Ald*, 158 kDa) are indicated by arrows. The polymerization activity, measured by the incorporation of [α-32P]dATP into salmon testes DNA, was based on the mean of two measurements with each fraction.>

<Fig. 5. Assay of primer extension and 3′–5′ exonuclease activity of the wild type and the mutants D1122A and D1124A. A, primer extension activity. One hundred nanograms of enzyme was used in each sample, and the reaction was carried out for 2 min at 70 °C; B, 3′–5′ exonuclease activity. Four hundred nanograms of enzyme was used, and the reaction was carried out for 10 min at 70 °C. Lanes 1, control without enzyme; lanes 2, wild type; lanes 3, D1122A; lanes 4, D1124A.>

(see “Materials and Methods”), and the DNA polymerization activities were measured and compared with a wild type protein. When the purified proteins were used, D1122A and D1124A showed only 0.51 and 0.82% activities of the wild type enzyme, respectively.

To further confirm the catalytic importance of Asp-1122 and Asp-1124, various mutants were made at these two residues, in which aspartic acid (D) was changed to asparagine (N), glutamic acid (E), glutamine (Q), and arginine (R), respectively. Their polymerization activities are shown in Fig. 6, A and B. D1122E and D1124E showed 15 and 43% of the activity of the wild type PolDPho. D1124N had 17% of the activity, although we were unable to obtain mutant D1122N protein. However, D1122Q, D1122R, D1124Q, and D1124R showed activities of less than 1% of the wild type. We also made two mutants at Gly-1123 (glycine to alanine or arginine) and measured their polymerization activities. It is interesting that G1123A had only 13% of the activity of the wild type enzyme, whereas G1123R had 0.8% activity. It is likely that the reduction is due to the proximity of Gly-1123 to Asp-1122 and Asp-1124.

SDS-PAGE analysis using purified mutant protein showed the same profiles as the wild type, indicating that the mutant proteins in the region were properly expressed and stable (data not shown). To know whether the activity loss of the Asp-1122, Gly-1123, and Asp-1124 mutants was caused by conformational changes, we measured the CD spectra of the mutant proteins. Except for D1122R, which had a CD profile rather different from the wild type protein (data not shown), the other mutants were identical (D1122A, D1122E, G1123A, G1123R, G1123R, D1124A, and D1124R).
D1124A, D1124E, D1124Q, and D1124R) or very similar (D1122E and D1122Q) to the wild type protein (Fig. 7, A–C).

The Mutants D1122A and D1124A Maintain the 3'–5' Exonuclease Activity—It is known that the polymerization and 3'–5' exonuclease domains are separate in family B DNA polymerase. If this is also the case in PolD, the mutants D1122A and D1124A should maintain the exonuclease activities. We checked the primer extension and 3'–5' exonuclease activities of both mutants by gel electrophoresis (Fig. 5). As expected, no DNA polymerization products from D1122A and D1124A could be detected (Fig. 5A); however, the 3'–5' exonuclease products were observed (Fig. 5B). We determined the reaction rates of exonuclease activities of the wild type enzyme, D1122A, and D1124A (Fig. 8). The reaction rates of the mutants D1122A and D1124A were 38 and 55% of the wild type, respectively. This shows that the exonuclease activities remain in the mutants, although they are slightly reduced. The data support the proposal that Asp-1122 and Asp-1124 are the catalytic residues of DNA polymerization and indicate that the catalytic residues of 3'–5' exonuclease and those of DNA polymerization are different in PolD.

The Mutants D1122E and D1124N Have Lower Magnesium Affinity than Does the Wild Type—We measured the magnesium requirements for polymerization activity of the mutants D1122E and D1124N. As shown in Fig. 9, the magnesium concentration required for D1122E to exhibit half of its own full activity was about 4.5 mM, and for D1124N it was more than 7.5 mM. Both are significantly higher than that of the wild type (about 2.5 mM). These facts may indicate that the affinities of D1122E and D1124N to magnesium became lower than the wild type protein and that the invariant residues Asp-1122 and Asp-1124 are probably involved in binding the magnesium that is important for polymerization activity.

DISCUSSION

In the present study, several advances were presented for the characterization of PolD. We identified two aspartate residues, Asp-1122 and Asp-1124, as the catalytic residues for DNA polymerization in PolD. They are located in the most

| Enzyme | Polymerization activity | 3'–5' exonuclease activity |
|--------|-------------------------|--------------------------|
| Wild type | +++ | ++ |
| E800A/E803A | + | ++ |
| D818A | +++ | +++ |
| D824A | +++ | +++ |
| D827A | +++ | +++ |
| E843A/E847A | +++ | +++ |
| D853A | + | ++ |
| E867A | + | ++ |
| D872A | +++ | +++ |
| D891A | +++ | +++ |
| D892A | +++ | +++ |
| E910A/D911A | +++ | +++ |
| D1122A | + | ND |
| D1122E | + | ND |
| D1122Q | + | ND |
| D1122R | + | ND |
| G1123A | + | ND |
| G1123R | + | ND |
| D1124A | + | ND |
| D1124E | + | ND |
| D1124Q | + | ND |
| D1124R | + | ND |
| E1125A | +++ | +++ |
| D1126A | +++ | +++ |
| D1133A | +++ | +++ |
| D1152A | +++ | +++ |
| D1162A | +++ | +++ |
| E1165A | +++ | +++ |
| D1167A | +++ | +++ |
| E1169A | +++ | +++ |
| D1174A | +++ | +++ |
| E1182A/E1185 | +++ | +++ |

*The underline indicates that purified proteins were used for the assay; the other samples used a heat-treated supernatant. 

Reactions were performed using 0.5 μg of supernatant or 0.2 μg of purified proteins by the acid precipitate method.

Reactions were performed using 0.5 μg of supernatant or 0.4 μg of purified proteins.

+++, ++, and +++++, indicate 1–20%, 21–60%, and 61–100% of wild type enzyme activity, respectively.

—, indicates undetectable or very low activity (<1% of the wild type enzyme).

ND, not determined.

FIG. 6. Comparison of polymerization activity of the wild type and the mutant PolDPho at the Asp-1122, Gly-1123, and Asp-1124 residues using the acid precipitate method. The activity was the mean of at least three time measurements for each protein and time. A, dNTP incorporation by enzymes measured after reaction for 30 min at 75 °C using 200 ng of proteins. The standard deviations are indicated at the top of the mean values. B, time courses of the polymerization activities using the wild type enzyme (●) and the mutant enzymes D1124E (■), D1122E (+), and D1124N (○). Two hundred nanograms of enzyme was used for each case.
conserved region of DP, YAHPY(F)FHA(S)AKRRCDG-(S)DED (Fig. 2). The change of Asp-1122 and Asp-1124 to alanine caused a dramatic decrease in polymerization activity. The CD analysis confirmed that the change in activity was not because of structural alteration. We presented evidence that Asp-1122 and Asp-1124 are probably involved in magnesium binding by analyzing the Mg\(^{2+}\) dependence for the mutants D1122E and D1124N. We further demonstrated that the mutants D1122A and D1124A maintained exonuclease activity, indicating that neither Asp-1122 nor Asp-1124 is the catalytic residue of the exonuclease activity.

Interestingly, the most conserved region is inserted by a mini-intein in three of the eight Archaea species with PolD, P. horikoshii, P. abyssi, and Halobacterium sp. NRC-1 but not others. Inteins tend to be inserted into important motifs for enzymatic activities in archaeal DNA replication proteins, such as family B DNA polymerase and replication factor C, as described by Cann and Ishino (3). This is indirect evidence that the most conserved motif in PolD has catalytic importance.

The catalytic sites of family B DNA polymerase have been located by mutagenesis (15). The metal binding residues are invariant aspartic acid residues in the most conserved motif -YGDTDS- in region I (or motif C). The catalytic residues of family A DNA polymerase have also been determined by mutagenesis and by structural analysis (16). An invariant aspartic acid...
acid residue in motif A (Asp-705 in \textit{E. coli} DNA polymerase I) and one in motif C (Asp-882 in \textit{E. coli} DNA polymerase I) are involved in metal binding. An additional carboxylate triad (Glu-710 and Glu-883) was recently found to be obligatory for catalysis in \textit{E. coli} DNA polymerase I (17). The sequences of family B and family A DNA polymerase in the catalytic regions are not conserved, but protein structural conservation has been proposed (15). In the superposition of family A on family B DNA polymerase, the pair of negatively charged residues, DE, of a $\beta$-$\beta$ turn in \textit{E. coli} DNA polymerase I matched the two negatively charged residues in human polymerase $\alpha$, which is a family B DNA polymerase (15). For PolD, sequences similar to motif A and motif C were proposed (3) according to visual inspection. However, mutagenesis at the proposed invariant residues in our study did not cause drastic changes in DNA polymerization. Instead, the invariant aspartic acid residues at Asp-1122 and Asp-1124 in the most conserved motif were found to be the catalytic sites. Therefore the most conserved motif is responsible for the polymerization catalysis. In this motif, glycine (or serine) is inserted between two aspartic acids. This is similar to the catalytic residues in motif C of family B DNA polymerase in which a threonine is inserted between two aspartic acid residues. Structurally, the threonine (T) was proposed to be at the bottom of two $\beta$ sheets. The glycine (or serine) with small side chains in the DP2 motif may perform the same functions as threonine in motif C of family B DNA polymerase. Structural analysis of PolDPho may eventually verify this hypothesis.

We also determined the molecular mass of PolDPho as 421 kDa. There are two possible combinations of subunit composition, in theory, based on this molecular mass; one is two large and two small subunits (predicted molecular mass of 430 kDa), and the other is four small subunits and one large subunit (predicted molecular mass of 428 kDa). The second combination does not seem likely considering the band intensity in the SDS-PAGE analysis (Fig. 3). Thus, we propose that PolDPho forms a heterotetrameric structure containing two large subunits and two small subunits in one molecule. Using gel filtration, we determined the molecular mass of the His-tagged small subunit to be around 70 kDa, corresponding to a monomer (data not shown). The small subunit could not form a tertiary structure; therefore the small subunit alone is not involved in the dimerization of the PolDPho complex. Further investigation is required to determine whether the large subunit alone or both subunits are involved in this process and whether the heterotetrameric structure exists \textit{in vivo}. The structure of recombinant PolDPho seemed to be different from that of recombinant PolD from \textit{P. furiosus}, which was reported to form a heterodimer (5). However, from the gel filtration results of the native PolD from \textit{P. furiosus}, an early elution of active protein occurred before the heterodimeric peak (4), and the specific activity of the early peak seemed to be higher than that of the major peak. We suppose that the early elution peak might be the heterotramer of the PolD from \textit{P. furiosus}.

The structures of the main DNA polymerases seem to be conserved in the three domains of life. It is known that the $\alpha$-$\epsilon$ subassembly of holoenzyme III of \textit{E. coli} is an asymmetric dimer including two $\alpha$ subunits that are responsible for the leading strand and lagging strand synthesis, respectively, of the DNA replication fork. In eukaryotes, DNA polymerase $\delta$ and $\epsilon$ are responsible for longer DNA synthesis in both strands after synthesis of short DNA fragments by DNA polymerase $\alpha$ (18). It has been demonstrated that yeast (\textit{Schizosaccharomyces pombe}) recombinant and native DNA polymerase $\delta$ complexes composed of four subunits form dimeric structures (19). There has also been evidence that budding yeast \textit{Saccharomyces cerevisiae} DNA polymerase $\epsilon$ complex forms a dimeric structure \textit{in vivo} and that the dimerization is essential for DNA replication (20). Therefore it is likely that the major DNA replication polymerases in the three domains of life are in a dimeric structure in general.

Finally, a co-expression system has been developed that has allowed us to produce sufficient amounts of proteins and to generate various mutant plasmids and enzymes. Several factors may contribute to the success of this system. One is the use of the His tag at the N terminus of the small subunit, which allows purification of the complex by pulling down the large subunit using the nickel column. The second is the use of an induction temperature of 30 °C rather than 37 °C and an isopropylthio-$\beta$-D-galactoside concentration of 2 mM. Perhaps the most important factor is the use of the \textit{E. coli} strain BL21-CondonPlus (DE3)-RI as the expression host. Previously, the strain BL21 (DE3) was used, but the expression level was extremely low. The improvement of protein expression by using strain BL21-CondonPlus (DE3)-RI is reasonable because DP1Pho and DP2Pho genes contain many rare codons for \textit{E. coli}. This co-expression system is also applicable for making other mutants in both subunits for further biochemical analysis. The system is now being evaluated for the commercial production of PolDPho.

Acknowledgments—We thank E. Yamamoto and H. Tokue for technical help during this study.

REFERENCES

1. Ito, J., and Braithwaite, D. K. (1991) \textit{Nucleic Acids Res.} 19, 4045–4057
2. Braithwaite, D. K., and Ito, J. (1993) \textit{Nucleic Acids Res.} 21, 787–802
3. Cann, I. K., and Ishino, Y. (1999) \textit{Genetics} 152, 1249–1267
4. Isamamura, M., Uemori, T., Kato, I., and Ishino, Y. (1995) \textit{Biol. Pharm. Bull.} 18, 1647–1652
5. Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, Y. (1997) \textit{Genes Cells} 2, 499–512
6. Ishino, Y., Komori, K., Cann, I. K., and Koga, Y. (1998) \textit{J. Bacterial.} 180, 2232–2236
7. Cann, I. K., Komori, K., Toh, H., Kanai, S., and Ishino, Y. (1998) \textit{Proc. Natl. Acad. Sci. U. S. A.} 95, 14250–14255
8. Miyakawa, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. (2000) \textit{Science} 288, 2212–2215
9. Edgell, D. R., and Donitle, W. F. (1997) \textit{Cell} 89, 965–988
10. Horton, R. M., Huht, H. D., Ho, S. N., Pullen J. K., and Pease L. R. (1989) \textit{Gene (Amst.)} 77, 61–68
11. Kadowaki, H., Kadowaki, T., Wandsford, E. F., and Taylor, S. I. (1989) \textit{Gene (Amst.)} 78, 161–166
12. Kimmel, T. A. (1985) \textit{J. Biol. Chem.} 260, 5787–5796
13. Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Szurka, T., Miyajima, N., Sugimura, M., and Tabata, S. (1995) \textit{DNA Res.} 2, 153–166
14. Perler, F. B. (1998) \textit{Col. 92, 1–4
15. Copeland, W. C., and Wang, T. S. (1993) \textit{J. Biol. Chem.} 268, 11028–11040
16. Steitz, T. A. (1998) \textit{Nature} 391, 231–232
17. Ganier, T. R., Kaushik, N., Singh, K., and Modak, M. J. (2000) \textit{J. Biol. Chem.} 275, 19685–19692
18. Bambara, R. A. (1997) \textit{J. Biol. Chem.} 272, 4647–4650
19. Zuo, S., Bermudez, V., Zhang, G., Kelman, Z., and Hurwitz, J. (2000) \textit{J. Biol. Chem.} 275, 5153–5162
20. Dua, R., Edwards, S., Levy, D. L., and Campbell, J. L. (2000) \textit{J. Biol. Chem.} 275, 28816–28825
