© 1996 Oxford University Press

The γ subfamily of DNA polymerases: cloning of a developmentally regulated cDNA encoding *Xenopus laevis* mitochondrial DNA polymerase γ

Fei Ye+, José A. Carrodeguas and Daniel F. Bogenhagen*

Department of Pharmacological Sciences, State University of New York at Stony Brook, Stony Brook, NY 11794-8651, USA

Received January 2, 1996; Revised and Accepted February 23, 1996

GenBank accession nos U49509 and U49510

ABSTRACT

We used the known sequence of the *Saccharomyces cerevisiae* DNA polymerase γ to clone the genes or cDNAs encoding this enzyme in two other yeasts, *Pychia pastoris* and *Schizosaccharomyces pombe*, and one higher eukaryote, *Xenopus laevis*. To confirm the identity of the final *X.laevis* clone, two antisera raised against peptide sequences were shown to react with DNA polymerase γ purified from *X.laevis* oocyte mitochondria. A developmentally regulated 4.6 kb mRNA is recognized on Northern blots of oocyte RNA using the *X.laevis* cDNA. Comparison of the four DNA polymerase γ gene sequences revealed several highly conserved sequence blocks, comprising an N-terminal 3′→5′ exonuclease domain and a C-terminal polymerase active center interspersed with γ-specific gene sequences. The consensus sequences for the DNA polymerase γ exonuclease and polymerase domains show extensive sequence similarity to DNA polymerase I from *Escherichia coli*. Sequence conservation is greatest for residues located near the active centers of the exo and pol domains of the *E.coli* DNA polymerase I structure. The domain separating the exonuclease and polymerase active sites is larger in DNA polymerase γ than in other members of family A (DNA polymerase I-like) polymerases. The *S.cerevisiae* DNA polymerase γ is atypical in that it includes a 240 residue C-terminal extension that is not found in the other members of the DNA polymerase γ family, or in other family A DNA polymerases.

INTRODUCTION

Our laboratory is interested in the structure and function of DNA polymerase γ, the enzyme responsible for replication of mitochondrial DNA. In this paper we report the molecular cloning of *Xenopus laevis* DNA polymerase γ cDNA using the sequence of the gene encoding DNA polymerase γ in the budding yeast *Saccharomyces cerevisiae* (1). Although *S.cerevisiae* and *X.laevis* are highly divergent species, the purified *X.laevis* DNA polymerase γ (2,3) shares a number of features with the yeast enzyme. Both enzymes efficiently utilize poly(rA)·oligo(dT) templates, respond similarly to polymerase inhibitors, possess associated 3′→5′ exonuclease activity and contain a DNA polymerase catalytic subunit of ~140 kDa. These features are shared with DNA polymerase γ purified from other sources as well (4–8), although the catalytic subunit of the *Drosophila* enzyme is somewhat smaller than that for vertebrates.

The primary sequence of *S.cerevisiae* DNA polymerase γ (1) shows similarity to those of other DNA polymerase I-like (family A) DNA polymerases in the polymerase and exonuclease domains (9,10). However, the *S.cerevisiae* DNA polymerase γ presents a rather poor match to the overall consensus sequence for family A polymerases in that it contains only one third of the highly conserved residues identified by Braithwaite and Ito (11). We first designed PCR primers using sequences in the pol domain to obtain the sequence of the gene from a second budding yeast, *Pychia pastoris*. This sequence information permitted the design of new primers to clone DNA polymerase γ from the fission yeast *Schizosaccharomyces pombe*. The homology among these three fungal DNA polymerase γ sequences enabled us to clone the cDNA for the *X.laevis* enzyme. In this paper we report these novel gene sequences and provide a comparison among the γ subfamily of DNA polymerases and other family A DNA polymerases.

MATERIALS AND METHODS

Yeast strains and growth media

Kluyveromyces lactis genomic DNA and *Pychia pastoris* strain SB 1099 were obtained from N. Dean (SUNY, Stony Brook, NY) and *Schizosaccharomyces pombe* strain SP1301 was obtained from R. Sternglanz (SUNY, Stony Brook, NY). The cells were either grown in YPAD broth (2% peptone, 1% yeast extract, 40 mg/ml adenine sulfate and 2% glucose) or YE (2.5% ethanol, 2% yeast extract) solid medium containing 1.5% agar.

*To whom correspondence should be addressed

+Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

PCR cloning

The PCR amplification reactions were performed using degenerate oligonucleotide primers to amplify sequences from 0.1 µg genomic DNA isolated from *P.pastoris*, *K.lactis* or *S.pombe* according to Moreno *et al*. (12). At sites of 2- or 3-fold degeneracy, mixed nucleotide residues were incorporated. At sites of 4-fold degeneracy, inosine was incorporated. For PCR with *P.pastoris* genomic DNA, amplification reactions were performed for 30 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 2 min under reaction conditions suggested by the manufacturer of *Taq* polymerase (Perkin Elmer-Cetus). The amplification reactions using *S.pombe* DNA were performed for three cycles of 94°C for 1 min, 37°C for 1 min and elongation at 72°C for 2 min, followed by an additional 30 cycles of 94°C for 1 min, 45°C for 1 min and 72°C for 2 min. Fragments of the expected size were gel purified and cloned into the pCRII vector (Invitrogen). Cloned PCR fragments were labeled by random priming and used to screen either a partial genomic *P.pastoris* library or λZap cDNA libraries (*S.pombe* library kindly provided by D. Beach; *X.laevis* library kindly provided by M. Roth). Hybridizations were continued for 16 h at 65°C in 5× SSC (1× SSC = 0.15 M NaCl, 0.015 M Na_3citrate), 5× Denhardt's solution, 0.5% SDS and 100 µg/ml calf thymus DNA. A final wash was performed in 1× SSC, 0.1% SDS at 65°C for 1 h.

A fragment of the *X.laevis* DNA polymerase γ cDNA was obtained by two cycles of nested PCR. In the first round the primers employed the sense strand of codons AVENTWLT and the antisense strand of codons WTRA(F+M)FCQ. The first round amplification reactions were performed for three cycles of 94°C for 1 min, 35°C for 1 min and 72°C for 2 min, followed by an additional 30 cycles of 94°C for 1 min, 45°C for 1 min and 72°C for 2 min. For the second round of amplification the degenerate primers represented the sense strand of codons FVGADVD and the antisense strand of codons AK(V+I)FNYGR, using the product of the first round reaction as the template.

The 5′-end of the *X.laevis* DNA polymerase γ was obtained by 5′-RACE (13) using *Pfu* polymerase (Stratagene) to reduce the potential for replication errors. Positive clones were sequenced using Sequenase (United States Biochemical Corp.) in the chain terminator (dideoxy) method using gene-specific primers.

RNA blot hybridization

Total RNA was prepared from defined numbers of staged *X.laevis* oocytes as described (14). RNA was resuspended in 10 mM Tris, pH 8, 1 mM EDTA, 0.5% SDS at a concentration of 1 µl/oocyte equivalent for oocyte stages I–III and 2 µl/oocyte for oocyte stages IV–VI. RNA was subjected to electrophoresis on a formaldehyde–agarose gel (15) and transferred to a Nytran membrane (Schleicher & Schuell). Hybridization with a probe prepared by random priming (16) of a 3.6 kb restriction fragment containing the entire coding region of the *X.laevis* DNA polymerase γ cDNA, excluding the signal sequence, was performed at 65°C in 6× SSC, 5× Denhardt's reagent, 100 µg/ml calf thymus DNA, 0.5% SDS. Filters were washed with 2× SSC, 0.5% SDS followed by 0.1× SSC/0.1% SDS at 65°C and hybridization was detected by autoradiography.

Immunological methods

Peptides corresponding to two conserved regions of the *X.laevis* DNA polymerase γ sequence, TRRAVEPTWLTASNC and KVFNYGRIYGAGC, were synthesized by Biosynthesis Inc., coupled to ovalbumin and injected into rabbits (nos 338 and 339 respectively) to prepare antisera using standard techniques (17). Antipeptide antisera were prepared by Cocalico Biologicals. A separate polyclonal antibody raised against purified 140 kDa *X.laevis* DNA polymerase γ by N. Insdorf of this laboratory was used as a positive control. These antisera and the pre-immune sera for the two antipeptide reagents were diluted 1:20 000 for use in immunoblotting experiments. Following overnight incubation with the primary antisera, filters were washed with phosphate-buffered saline containing 0.5% Tween 20 and incubated with a 1:5000 dilution of calf alkaline phosphatase-conjugated goat anti-rabbit antibodies (KPL Laboratories). Following additional washing of the membrane, proteins were identified with colorimetric detection reagents (KPL Laboratories).

RESULTS

Cloning of DNA polymerase γ from *P.pastoris* and *S.pombe*

In our first attempts to use the sequences of *S.cerevisiae* DNA polymerase γ to clone a second DNA polymerase γ we performed PCR reactions with genomic DNA derived from both *K.lactis* and *P.pastoris.* Two amino acid sequences conserved among the polymerase domains of the *S.cerevisiae* DNA polymerase γ gene and polymerase 1-like DNA polymerase genes, FNYGRIY and SIHDEIR, were used to design degenerate oligonucleotides. Fragments of the expected size were obtained from both amplification reactions. Since the *K.lactis* sequence was found to be very closely related to that of the *S.cerevisiae* DNA polymerase γ, further analysis of this gene was not pursued. The *P.pastoris* PCR product was used to isolate a clone containing a 4 kb *Eco*RI insert of genomic DNA. This insert was completely sequenced. The sequence was found to lack both an initiator methionine and homology to the extreme N-terminus of *S.cerevisiae* DNA polymerase γ. An overlapping clone was isolated and the N-terminus of DNA polymerase γ was sequenced. The composite sequence of the *P.pastoris* DNA polymerase γ gene has been submitted to GenBank (accession no. U49510).

Using the homology between DNA polymerase γ from *S.cerevisiae* and *P.pastoris*, degenerate PCR primers were designed for the peptide sequences YCFVGAD and NWAIQS. These degenerate PCR primers were used with *S.pombe* genomic DNA to amplify a 548 bp product that was cloned into the pCR II vector and sequenced. The sequence revealed 62.2 and 59.8% nucleotide sequence homology and 62.4 and 61.4% amino acid identity to the *S.cerevisiae* and *P.pastoris* DNA polymerase γ genes and proteins respectively. This fragment was used as a probe for screening 400 000 recombinant phage plaques from an *S.pombe* cDNA library (gift from D. Beach, Cold Spring Harbor Laboratory, cloned in λZapII; Stratagene, La Jolla, CA). Ten positive λ clones were isolated and the DNA inserts were recovered by *in vivo* excision following superinfection with an M13 helper phage. The largest cDNA clone, which spans 3545 bp, was fully sequenced with Sequenase using gene-specific primers. Beginning at residue 286 there is an ATG codon followed by an open reading frame of 1018 amino acids. Southern blotting experiments confirmed that both *S.pombe* and *P.pastoris*

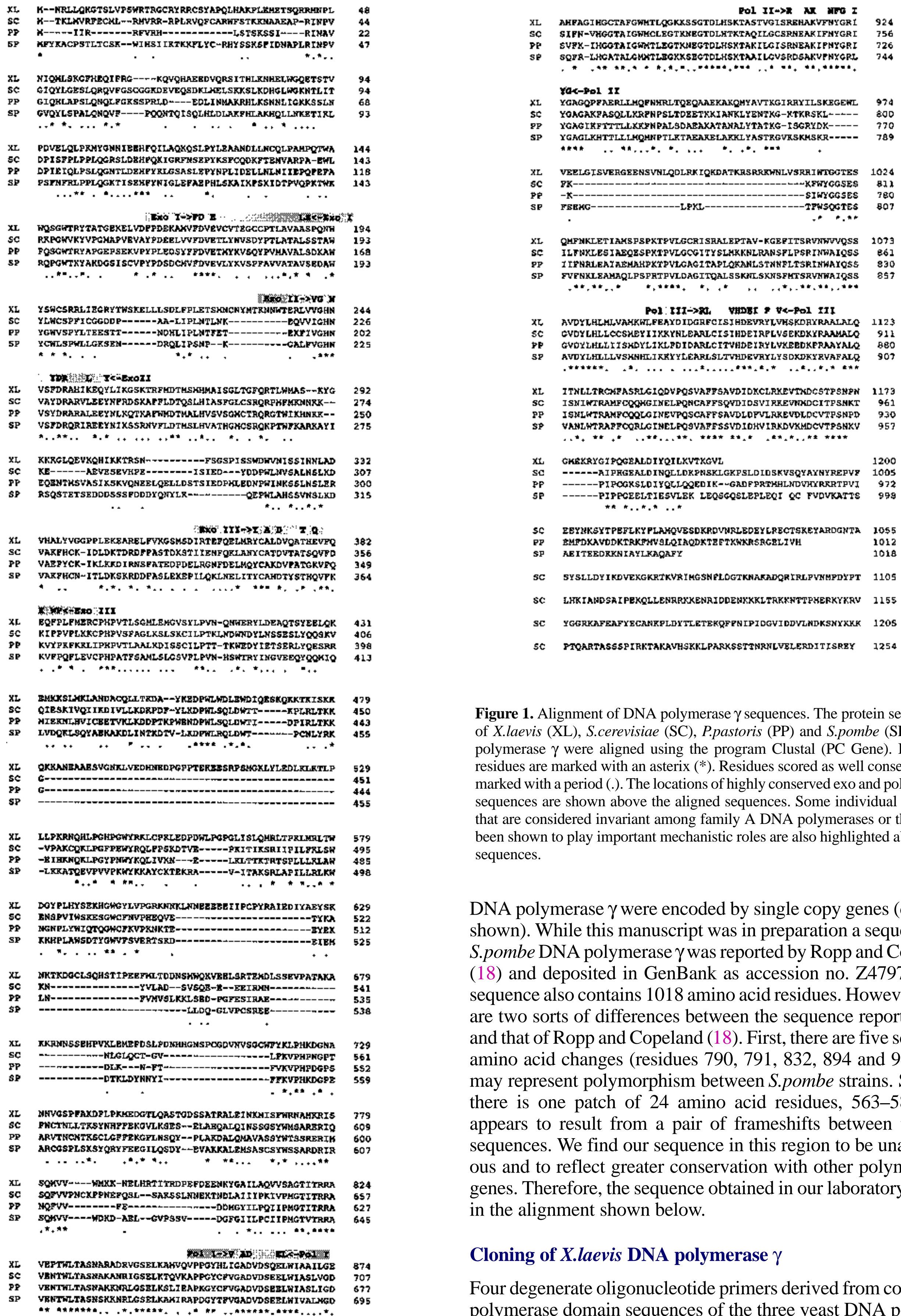

Figure 1. Alignment of DNA polymerase γ sequences. The protein sequences of *X.laevis* (XL), *S.cerevisiae* (SC), *P.pastoris* (PP) and *S.pombe* (SP) DNA polymerase γ were aligned using the program Clustal (PC Gene). Identical residues are marked with an asterix (*). Residues scored as well conserved are marked with a period (.). The locations of highly conserved exo and pol domain sequences are shown above the aligned sequences. Some individual residues that are considered invariant among family A DNA polymerases or that have been shown to play important mechanistic roles are also highlighted above the sequences.

DNA polymerase γ were encoded by single copy genes (data not shown). While this manuscript was in preparation a sequence for *S.pombe* DNA polymerase γ was reported by Ropp and Copeland (18) and deposited in GenBank as accession no. Z47976. This sequence also contains 1018 amino acid residues. However, there are two sorts of differences between the sequence reported here and that of Ropp and Copeland (18). First, there are five scattered amino acid changes (residues 790, 791, 832, 894 and 942) that may represent polymorphism between *S.pombe* strains. Second, there is one patch of 24 amino acid residues, 563–586, that appears to result from a pair of frameshifts between the two sequences. We find our sequence in this region to be unambiguous and to reflect greater conservation with other polymerase γ genes. Therefore, the sequence obtained in our laboratory is used in the alignment shown below.

Cloning of *X.laevis* DNA polymerase γ

Four degenerate oligonucleotide primers derived from conserved polymerase domain sequences of the three yeast DNA polymer-

ase γ genes were used to amplify fragments of *X.laevis* cDNA by two rounds of nested PCR as described in Materials and Methods. This provided a 195 bp product that was cloned into the pCR II vector and sequenced. The sequence showed 60–65% amino acid identity to each of the sequences of the yeast DNA polymerase γ genes. This fragment was used as a probe for screening 800 000 recombinant phage plaques from an *X.laevis* ovary cDNA library (gift of M. Roth, Hutchinson Cancer Center, Seattle, WA). Hybridizations were performed as described above. Eleven positive λ clones were isolated and plaque purified and the DNA inserts with the vector sequence carried by the λ clones were recovered by *in vivo* excision. One cDNA clone was fully sequenced using gene-specific primers. This clone contained an open reading frame of 663 amino acids with clear homology to the C-terminus of the yeast DNA polymerase γ genes. This was considered to be an incomplete cDNA clone, since it lacked an N-terminal methionine and sequences related to the N-terminal half of yeast DNA polymerase γ. The same 5′-end was observed in four independent λ isolates from this library, while other isolates were even more severely truncated. PCR reactions showed that longer isolates were not available in this library or in three other *X.laevis* oocyte and embryo cDNA libraries.

5′-RACE using *Pfu* polymerase was used to complete the cloning of the *X.laevis* DNA polymerase γ (13). The composite sequence of the overlapping clones contains 1200 amino acids with a predicted molecular mass of 137.3 kDa, in excellent agreement with the estimated molecular mass of the purified catalytic subunit (2). The DNA sequence (data not shown; deposited in GenBank as accession no. U49509) includes 212 residues of putative 5′ untranslated sequence, including two stop codons in-frame with the initiator ATG. The extreme N-terminus represents a potential mitochondrial transit sequence. These features suggest that the sequence contains the full extent of the open reading frame encoding DNA polymerase γ. A comparison of the protein sequence with those of other members of the polymerase γ family is shown in Figure 1 and discussed in more detail below.

To confirm that the isolated *X.laevis* cDNA encodes DNA polymerase γ antisera were raised against two conserved peptide sequences from the *X.laevis* cDNA sequence as described in Materials and Methods. These two antisera, 338 and 339, were used to probe a Western blot containing DNA polymerase γ purified from oocyte mitochondria. Both antisera recognized the same ~140 kDa DNA polymerase γ polypeptide identified by an antisera raised against the intact *X.laevis* polypeptide (Fig. 2). The blots shown in Figure 2 indicate that the pre-immune serum for 339 appeared to show a relatively high background of reactivity. All three antisera also recognized a polypeptide of ~100 kDa that appears to be a proteolytic fragment of the catalytic subunit. These results support the conclusion that the cloned gene does encode *X.laevis* DNA polymerase γ.

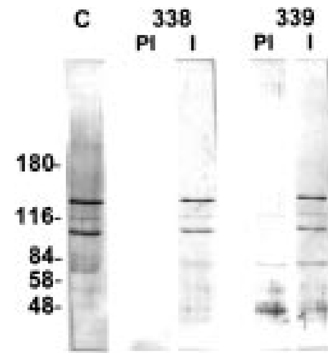

Figure 2. Antisera raised against peptide sequences of the putative *X.laevis* polymerase γ cDNA react with purified *X.laevis* DNA polymerase γ. A set of immunoblots is shown to test the ability of antipeptide antisera 338 and 339 (see Materials and Methods) to react with purified DNA polymerase γ. DNA polymerase γ was purified by a modification (31) of the procedure established by Insdorf and Bogenhagen (2). An amount of DNA polymerase γ sufficient to incorporate 1 nmol nucleotide in a standard reaction using poly(rA)·oligo(dT) template–primer was loaded in each of five lanes of a 6% PAGE–SDS gel. Following electrophoresis and transfer to an Immobilon membrane filter strips were incubated with 1:20 000 dilutions of a polyclonal antiserum raised against the gel-purified catalytic subunit of *X.laevis* DNA polymerase γ (lane C) or pre-immune (PI) or immune (I) sera obtained from rabbits 338 and 339 injected with the two synthetic peptides. The mobilities of pre-stained molecular weight standards are indicated on the left as M_r in kDa determined by the manufacturer (Sigma Chemical Co.).

DNA polymerase γ mRNA levels are developmentally regulated in *X.laevis* oocytes

RNA blot hybridization was performed to confirm the size and expression of the *X.laevis* DNA polymerase γ. Total RNA was isolated from defined numbers of oocytes and quantities of RNA equivalent to the contents of 10 immature oocytes or five stage IV–VI oocytes were subjected to electrophoresis on a formaldehyde–agarose gel. Lower numbers of cells were used for the mature oocytes due to the large quantity of rRNA accumulated at these stages. The RNA was transferred to a membrane and hybridized to a radioactive DNA probe containing the entire coding region of the DNA polymerase γ except for the signal sequence. An mRNA of ~4.6 kb was identified (Fig. 3), in excellent agreement with the 4396 nt cDNA sequence, which does not include a poly(A) tail. In this and other RNA blot experiments hybridization was most intense with stage II oocytes. This result is consistent with other experiments we have performed using internally standardized RT–PCR methods that show that absolute mRNA levels are decreased several-fold in mature oocytes (J. Carrodeguas, unpublished observations).

DISCUSSION

We used the sequence of *S.cerevisiae* DNA polymerase γ to clone the homologous genes or cDNAs from two other yeasts. The resulting sequence relationships permitted cloning and sequencing of the cDNA for *X.laevis* DNA polymerase γ. Confirmation that the final eukaryotic gene does indeed encode DNA polymerase γ was provided by the finding that two antisera raised against conserved sequences reacted with the purified *X.laevis* DNA polymerase γ (Fig. 2). Hybridization of the cDNA to total RNA isolated from staged oocytes revealed a 4.6 kb mRNA that declines in abundance as oocytes progress through vitellogenesis (Fig. 3). This result is also consistent with the well-established pattern of accumulation of mtDNA during oogenesis in *X.laevis*. Most of the ~100 000-fold amplification of mtDNA that occurs during oogenesis is complete by stage IV (19) and mtDNA replication is relatively quiescent in mature oocytes and early embryos (20). A more complete analysis of the levels of DNA polymerase γ protein and mRNA levels throughout oogenesis and embryogenesis is in progress and will be presented elsewhere.

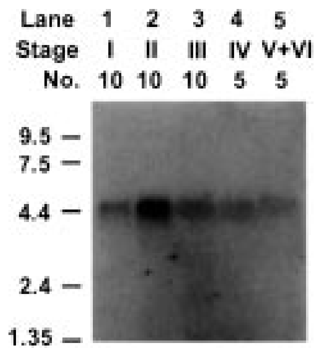

Figure 3. *Xenopus laevis* DNA polymerase γ mRNA levels are developmentally regulated during oogenesis. An autoradiogram of an RNA blot analysis is shown. RNA obtained from 10 stage I, II or III oocytes or from five stage IV or combined stage V+VI oocytes was subjected to electrophoresis on a 0.7% agarose gel containing formaldehyde and transferred to a charged nylon membrane. DNA polymerase γ mRNA was detected by hybridization to a ^{32}P probe prepared by random primed DNA synthesis using the entire *X.laevis* DNA polymerase γ cDNA. Relative mobilities of RNA standards (Life Technologies Inc.) are shown on the left in kb.

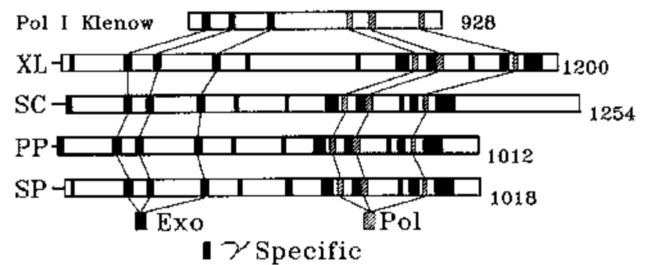

Figure 4. Diagram of the arrangement of conserved sequences in DNA polymerase γ. The locations of conserved exo and pol domain sequences are shown for the four DNA polymerase γ sequences (XL, *X.laevis*; SC, *S.cerevisiae*; PP, *P.pastoris*; SP, *S.pombe*) interspersed with sequence blocks conserved in the DNA polymerase γ subfamily (γ-specific). For comparison, the locations of conserved exo and pol sequences in the Klenow fragment (residues 324–928) of DNA polymerase I are also shown.

The sequence homology among members of the DNA polymerase γ subfamily is rather low. Pairwise comparison of sequences typically shows only 40–60% identity. However, alignment of all four genes (Fig. 1) reveals an ordered pattern of conserved sequences, as diagramed in Figure 4. All four polymerases contain N-terminal sequences recognizable as mitochondrial targeting signals, despite the fact that the primary sequence of these signals is not conserved. Some of these conserved sequence blocks comprise the 3′→5′ exonuclease and polymerase active centers, conserved in family A DNA polymerases, while others appear to be specific to the γ subfamily. The observation that both polymerase and 3′→5′ exonuclease domains are conserved in a vertebrate DNA polymerase γ suggests that the report that the 3′→5′ exonuclease domain of porcine DNA polymerase γ may reside in a separate subunit (8) is unlikely to be correct. It is quite possible that an N-terminal proteolytic fragment of DNA polymerase γ that retained exonuclease activity may have been present in the enzyme preparation used by Longley and Mosbaugh (8). Figures 1 and 4 show that the *X.laevis* and *S.cerevisiae* DNA polymerase γ sequences are appreciably longer than those from *P.pastoris* or *S.pombe*. The larger size of the *S.cerevisiae* enzyme was documented in the original cloning (1), while the larger size of the *X.laevis* DNA sequence is consistent with the previous report of the size of its catalytic subunit (2). The other two polymerases have not been studied previously. It is interesting to note that different patterns of sequence insertions or additions create the larger sizes of the *X.laevis* and *S.cerevisiae* polymerases. The *S.cerevisiae* enzyme contains a long C-terminal tail not found in the other DNA polymerase γ sequences, or in other members of family A DNA polymerases (11). The larger size of the *X.laevis* enzyme results mainly from a series of at least six substantial insertions (depending on the detailed sequence alignment) between the exo and pol domains.

We were interested in comparing DNA polymerase γ sequences with that of DNA polymerase I of *E.coli*, the prototype of family A DNA polymerases. This comparison is of particular interest since the structure of the Klenow fragment of DNA polymerase I has been determined (21). More recent refinements of this structure (22–25) and analysis of the structure of HIV reverse transcriptase (26) have made a remarkable contribution to understanding the common structure of DNA polymerases (27). By visual inspection of the DNA polymerase γ sequences it was apparent that homology with the DNA polymerase I large fragment was not limited strictly to the Exo I, II and III or Pol I, II and III blocks identified by Ito and Braithwaite (10) and Blanco *et al.* (9). Therefore, to provide an optimal alignment the sequences of the 3′→5′ exonuclease and polymerase domains were used to generate a consensus sequence for γ polymerases. A comparison of this consensus sequence with the relevant portions of the DNA polymerase I large fragment is shown in Figure 5. The high degree of primary sequence conservation between γ polymerases and the Klenow fragment suggests that aspects of the three-dimensional structure will be conserved as well. Therefore, amino acid residues in the Klenow fragment that are conserved in the consensus DNA polymerase γ sequence (asterisks in Fig. 5) are illustrated within the structural model for the Klenow fragment in Figure 6.

We will consider the exo domain first. Most of the critically important residues in the exo domain of the fragment identified by Beese and Steitz (23) and mutagenized by Derbyshire *et al.* (28) are conserved in DNA polymerase γ sequences. These include T358 (V in some sequences), Y497 and E357, which are thought to interact with the dTMP product of the exonuclease. The residues that interact with the two metal ion cofactors are also conserved: D355, D424, D501 and, again, E357. Residues interacting with the second and third bases of the single-strand DNA in the exo domain, N420, Y431 and M443 (Q in *S.cerevisiae*) are highly conserved in DNA polymerase γ. There are two exceptions to this pattern of sequence conservation. The DNA polymerase γ sequences do not appear to contain DNA polymerase I residues L361 and F473, which are thought to participate in non-sequence-specific hydrophobic interactions with the DNA base to be excised. It is interesting that γ polymerases appear to have an Ile residue replacing F473 (Fig.5). However, DNA polymerase γ sequences contain insertions of 32–49 residues at a site that would separate helix D and helix E of DNA polymerase I that make it difficult to identify a residue that might substitute for L361. Three of the highly conserved residues (corresponding to Klenow residues D355, D424 and D501) have been mutagenized in *S.cerevisiae* DNA polymerase γ by Foury and Vanderstraeten (29) and shown to be required for

A. Exo Domain

```
                      -------Exo I-----
GAM CON           V.YPDED.MVFDVETLY.VS.YPTLAVALS-------AWY.W.SP.L..............
DNA pol I 344     AKLEKAPVFAFDTETDSLDNISANL-VGLSFAIEPGVAAYIPVAHDYLDAFDQISRERALE  403
                           .**.**       ..* *.**      *  *

                               ----Exo II-----
GAM CON           LLIPLNT......VGHNVSYDRARILEEYNIK.SKA.FMDTMSLHVA..GLCSRQR..-
DNA pol I 404     LLKPLLEDEKALKVGQNLKYDRG.ILANYGI-ELRGIAFDTMLESYILNSVAGRHDMDS 460
                   * **.       **.*. ***. **. * *  ... .***         .. *

                                                      ----Exo III---
GAM CON           -.EPWLN.SSLNSL.DVAFCILDK..RD.FAS.D-IFL....YCA.DV.AT.QVF.KVFP.FL.
DNA pol I 461     LAERWLKHKTI.TFEEIAG----KG-KNQLTFNQIALEEAGRYAAEDADVTLQLHLKMWPDLQK 518
                    *  *   .. .   .*     *  .              *.*  . . * *  *..*
```

B. Pol Domain

```
GAM CON           GAIL...IVPMGTITRRAVENTWLTASNAK.NRLGSELK.MV.AP.GYCF
DNA pol I 667     G-RLSSTDPNLQNIPVRNEE--------------GRRIRQAFIAPEDYVI   701
                  *  *        .*. *.  *              *.  ..    ** .*  .

                  ----Pol I-----
GAM CON           VGADVDSEELWIAALLGDS.F..IHGGTAIGWNTLEG.K.EGTDLHSKTA
DNA pol I 702     VSADYSQIELRIMAHLSRD--KGLLTAFAEGKDIHRATAAEVFGLPLET-   748
                  *.** .  **.* * *. .    . .  * *     .. .*  .*. .*

                        ----Pol II-----
GAM CON           .ILGISR..AKVFNYGRIYGAG.KFAT.LL..FNP.LT..EAK..A.LY.
DNA pol I 749     -VTSEQRRSAKAINFGLIYGMSA-FG--LARQLNIPRK--EAQKYMDLYF   792
                   . .  *  ** .*.* *** . *.   *.. .*      **    .**

GAM CON           ..TKG..S....-W.GGSES..FNKLEAIA...SPKTPVLGAGIT..L...N
DNA pol I 793     ERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDIKSSNGARR-------- 836
                      *                    *              **

                                                           --Pol III
GAM CON           L..N.FLTSRVNWAIQSSAVDYLHLLLVSM.YLIK.Y.IDARLCISIHDE
DNA pol I 837     AAAERAAINAPM---QGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDE   883
                                 *..*.* ..    ...  ..      *. . .***

                  ---------
GAM CON           VRYLV.EKDKYRAA.ALQISNLWTRAMFCQRLGINELPQSVAFFS.VDID.
DNA pol I 884     LVFEVHKDDV--DAVAKQIHQLMEN---CTRLDVPLLVE-VGSGENWDQAH   928
                  . . * ..*    * * ** .. ..     **..  * . *.     *
```

Figure 5. Comparison of consensus sequences of the exo and pol domains of DNA polymerase γ with the corresponding region of DNA polymerase I (Klenow fragment) of *E.coli*. Residues were scored as consensus residues if they were found in three of the four DNA polymerase γ sequences or if they were found in two sequences with chemically similar residues in the other DNA polymerase γ sequences. The resulting sequence (GamCon) was aligned with the relevant sections of the Klenow fragment of *E.coli* DNA polymerase I. Dashes (–) are used to indicate gaps in the sequences required for optimal alignment. Periods (.) are used to indicate single amino acid residues in the Gam Con sequence that are not conserved among the four DNA polymerase γ sequences. (**A**) Sequence comparisons in the 3′→5′ exonuclease domain show the locations of conserved sequences noted by Blanco *et al*. (9). Note that the residue in the GamCon sequence aligned to DNA polymerase I residue 358 is V in two of the sequences listed in Figure 1. (**B**) Sequence comparisons in the pol domain show the locations of conserved sequences noted by Ito and Braithwaite (10).

exonuclease activity. Figure 6 shows that there is a strong tendency for identical residues to be located in a central core of the exo domain, with less sequence conservation mapping in outlying regions. It is interesting to note that the conserved sequences in the exo II motif tend to lie along one face of helix C in the DNA polymerase I large fragment structure (Fig. 6). Similarly, the identical residues in the exo III motif principally lie along one face of the center of the long helix F.

A similar arrangement of conserved sequences is seen in the pol domain. Two groups have previously noted the high degree of sequence similarity between polymerase active site residues of *S.cerevisiae* DNA polymerase γ and family A DNA polymerases (9,10). These sequence blocks, which are referred to as Pol I, II and III domains by Ito and Braithwaite (10), are also highly conserved in the other members of the DNA polymerase γ family (Figs 1 and 5). These sequences include amino acid residues identified as critical elements of the active site (Pol block II) and the nucleotide binding domain (Pol block III). Figure 6 shows that these conserved residues are dispersed in the regions of the pol domain that have been compared to the 'palm' and 'fingers' of a right hand structure (26,27). In contrast, there is little sequence conservation in helix G, which connects the exo and pol domains of DNA polymerase I, or in helices H and I, which form the 'thumb' of the polymerase structure. DNA polymerase I contains 185 amino acid residues separating the 'Exo III' and the 'Pol I' blocks. The corresponding regions of DNA polymerase γ sequences are much larger: 479 and 338 residues separate these two motifs in *X.laevis* and *S.cerevisiae* DNA polymerase γ respectively. It is interesting to note that this region of the *X.laevis* DNA polymerase γ contains a potential leucine zipper between residues 563 and 584. No specific protein partners for the catalytic domain of *X.laevis* DNA polymerase γ have been identified.

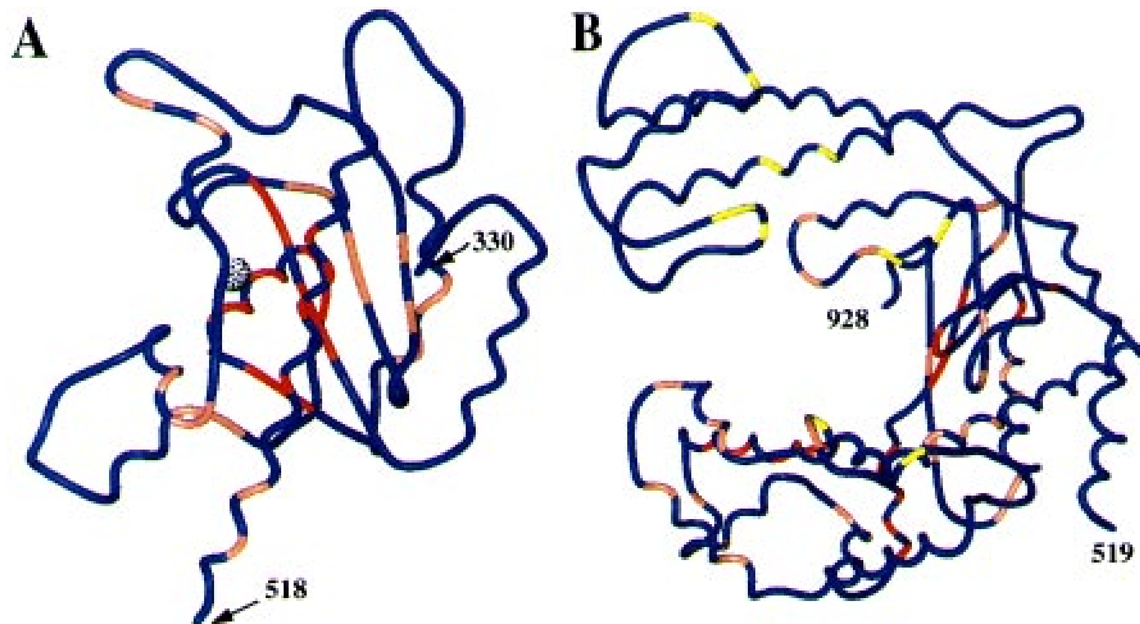

Figure 6. Locations of residues conserved in DNA polymerase γ within the structural model for the large fragment of DNA polymerase I. Ribbon diagrams of the exo domain (residues 330–518; **A**) and pol domain (residues 519–928; **B**) of the large fragment of DNA polymerase I were generated on a Silicon Graphics workstation using the Midas program. The Zn in the exo domain is shown as a black Van der Walls radius. Residues conserved in the consensus sequence for DNA polymerase γ are colored red for residues in the Exo I, II and III and Pol I, II and III motifs and pink for conserved residues outside these sequence blocks. In (B) yellow segments indicate DNA binding residues identified by Beese *et al*. (25) that are *not* conserved in DNA polymerase γ.

The segment of DNA polymerase I connecting the exo and pol domains includes a portion of the enzyme that was not well structured in the original model for the large fragment of DNA polymerase I (21). However, when the protein was co-crystallized with duplex DNA Beese *et al*. (25) found that much of this so-called 'disordered domain' was better defined and, in combination with helices H and I, formed a major part of the 'thumb' region that contributes to the groove in the enzyme that binds duplex DNA. Beese *et al*. (25) identified several residues from N579 to N678 that interacted non-specifically with the backbone phosphates of the DNA template and primer in this region that were highly conserved in polymerase I-like DNA polymerases. The alignment of polymerase I-like sequences by Braithwaite and Ito (11) did not identify these otherwise invariant residues in *S.cerevisiae* DNA polymerase γ. Computer analysis of the additional DNA polymerase γ sequences presented in this paper has also failed to identify these residues. As noted above, the connecting domain and putative thumb region of DNA polymerase γ is much larger than its counterpart in DNA polymerase I. Given the large difference in primary structure compared with DNA polymerase I, it is not feasible to identify individual DNA contact residues in DNA polymerase γ without direct structural analysis. We speculate that the large size of the domain separating the exo and pol active sites in DNA polymerase γ may allow this enzyme to enclose the DNA primer–template more tightly and may contribute to the high degree of processivity observed for the mitochondrial polymerase (30).

The alignment of DNA polymerase γ sequences (Fig. 1) also reveals a number of conserved sequence blocks that are not contained in other family A DNA polymerases (γ-specific blocks in Fig. 4). These conserved sequences are part of the structural scaffold required to position the exo and pol domain sequences. Some of these conserved sequences may also be required for specific interactions with other components of the mtDNA replication machinery. Although the functional roles of these γ-specific sequences are completely unknown, it is reasonable to expect that they should be found in the enzyme from other sources as well.

ACKNOWLEDGEMENTS

We thank D. Beach, N. Dean, F. Foury, M. Roth and R. Sternglanz for reagents used in this work and P.Fisher for review of the manuscript. This research was supported by grant NIH GM 29681.

REFERENCES

1 Foury,F. (1989) *J. Biol. Chem*., **264**, 20552–20560.
2 Insdorf,N.F. and Bogenhagen,D.F. (1989a) *J. Biol. Chem*., **264**, 21491–21497.
3 Insdorf,N.F. and Bogenhagen,D.F. (1989b) *J. Biol. Chem*., **264**, 21498–21503.
4 Wernette,C.M. and Kaguni,L.S. (1986) *J. Biol. Chem*., **261**, 14764–14770.
5 Olson,M.W. and Kaguni,L.S. (1992) *J. Biol. Chem*., **267**, 23136–23142.
6 Gray,H. and Wong,T.W. (1992) *J. Biol. Chem*., **267**, 5835–5841.
7 Kunkel,T.A. and Soni,A. (1988) *J. Biol. Chem*., **263**, 4450–4459.
8 Longley,M.J. and Mosbaugh,D.W. (1991) *J. Biol. Chem*., **266**, 24702–24711.
9 Blanco,L., Bernad,A. and Salas,M. (1991) *Nucleic Acids Res*., **19**, 955.
10 Ito,J. and Braithwaite,D. (1990) *Nucleic Acids Res*., **18**, 6716.
11 Braithwaite,D. and Ito,J. (1993) *Nucleic Acids Res*., **21**, 787–802.
12 Moreno,S., Klar,A. and Nurse,P. (1991) *Methods Enzymol*., **194**, 795–823.
13 Frohman,M. (1993) *Methods Enzymol*., **218**, 340–56
14 Chomczynski,P. and Sacchi,N. (1986) *Anal. Biochem*., **162**, 156–159
15 Rave,N., Crkvenjakov,R. and Boedtker,H. (1979) *Nucleic Acids Res*., **6**, 3559–3567

16 Feinberg,A. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13
17 Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
18 Ropp,P.A. and Copeland,W.C. (1995) *Gene*, **165**, 103–107.
19 Webb,A.C. and Smith,L.D. (1977) *Dev. Biol.*, **56**, 219–225.
20 Chase,J.W. and Dawid,I.B. (1972) *Dev. Biol.* 27, 504–518
21 Ollis,D., Brick,P., Hamlin,R., Xuong,N. and Steitz,T. (1985) *Nature*, **313**, 762–766.
22 Freemont,P., Friedman,J., Beese,L., Sanderson,M. and Steitz,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8924–8928.
23 Beese,L.S. and Steitz,T.A. (1991) *EMBO J.*, **10**, 25–33.
24 Beese,L.S., Friedman,J.M. and Steitz,T.A. (1993) *Biochemistry*, **32**, 14095–14101.
25 Beese,L.S., Derbyshire,V. and Steitz,T.A. (1993) *Science*, **260**, 352–355.
26 Kohlstaedt,L., Wang,J., Friedman,J., Rice,P. and Steitz,T. (1992) *Science*, **256**, 1783–1790.
27 Joyce,C. and Steitz,T. (1994) *Annu. Rev. Biochem.*, **63**, 777–822.
28 Derbyshire,V., Grindley,N. and Joyce,C. (1991) *EMBO J.*, **10**, 17–24.
29 Foury,F. and Vanderstraeten,S. (1992) *EMBO J.*, **11**, 2717–2726.
30 Williams,A.J., Wernette,C.M. and Kaguni,L.S. (1993) *J. Biol. Chem.*, **268**, 24855–24862.
31 Pinz,K.G., Shibutani,S. and Bogenhagen,D.F. (1995) *J. Biol. Chem.*, **270**, 9202–9206