Vol. 136, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1986 Pages 341-347

SEQUENCE OF HUMAN DNA POLYMERASE β mRNA OBTAINED THROUGH cDNA CLONING

D.N. SenGupta, B.Z. Zmudzka, P. Kumar, F. Cobianchi, J. Skowronski and S.H. Wilson

Laboratory of Biochemistry, National Cancer Institute
National Institutes of Health, Bethesda, MD 20892

Received March 6, 1986

Summary: A cDNA library from polyA$^+$ RNA of a human teratocarcinoma cell line in phage λgt11 was screened with a fragment of the rat β-polymerase cDNA, λpolβ-10, as probe. Five positive phage were identified and plaque purified. The cDNA of one positive clone selected for detailed study was 1257 bp. This insert was sequenced and found to contain the coding region for β-polymerase, as well as 163 bp and 137 bp from the 5' and 3' untranslated regions, respectively. The primary structure of human β-polymerase (318 amino acids, M_r=36,133) deduced from the cDNA was similar to rat β-polymerase (95% matched residues). The greatest difference between the sequences of the human and rat cDNAs was in the 3' untranslated regions (64% matched base residues). These results provide necessary sequence information for study of the human β-polymerase gene.

DNA polymerase β is considered to be one of the required proteins in DNA repair pathways of mammalian cells (1-3). Several human hereditary disorders, such as ataxia telangectasia and xeroderma pigmentosum, appear to be associated with defects in DNA repair (4), and it is of interest to examine β-polymerase genetics and expression in these and in other human disorders potentially involving defective DNA synthesis. To this end, we now have isolated a full-length cDNA for β-polymerase from a human library in λgt11.

A cDNA for newborn rat brain β-polymerase was recently cloned from a λgt11 library using a polyclonal antibody (5). A fragment of this sequence was used in Southern analysis of human-rodent hybrids to localize the single-copy human gene for β-polymerase to chromosome 8 (6). In addition, a restriction fragment length polymorphism was found using a probe from the 5' end of the cDNA. The human cDNA described here has provided sequence information and ideal probes for defining further polymorphisms in and around the β-polymerase gene, as well as studies of the structure and expression of the human gene.

Materials and Methods

Preparation of human cDNA library in λgt11. Cytoplasmic RNA from high density cultures of the NTera2D1 human teratocarcinoma cell line (7) was isolated and enriched for polyadenylated RNA by chromatography on a poly-U

0006-291X/86 $1.50

Sephadex (BRL) column as described (11). 2 µg of polyA$^+$ RNA was denatured in 10 mM methylmercury hydroxide and added to 100 µl of M-MLV reaction buffer (8) with the following modifications: oligo (dT)12-18 (PLBiochemicals) was at 100 µg/ml; 2-mercapoethanol and RNAsin (Promega Biotech) were included at concentrations of 7 mM and 500 U/ml, respectively; actinomycin D was omited. 400 U of M-MLV reverse transcriptase (BRL) were added and the reaction was incubated for 60 min. at 37°C. Then 400 µl of 50 mM Tris-HCl, pH 7.5, 75 mM KCl, 7 mM $MgCl_2$ were added and the second cDNA strand was synthesized as described (9). For subsequent steps of methylation of internal EcoRI sites, end-repair, attaching EcoRI linkers, size fractionation and ligation to the EcoRI cleaved lambda λgt11 vector (Vector Cloning Systems) published procedures were followed (10). Recombinant phage DNA was packaged using Gigapack packaging extract (Vector Cloning Systems) as recommended by the manufacturer. The final library contained about $5x10^6$ recombinants as judged to the phage phenotype (10).

Screening for human β-polymerase recombinant clones. The λgt11 cDNA library prepared with polyA$^+$ RNA from growing human teratocarcinoma cells, NTera2D1 (11), was used. Screening for DNA polymerase β clones was performed by the method described by Benton and Davis (12), except that 2 x 10^6 phage were cultured on each 150 mm plate for 4 h at 37°C. Duplicate filters were made from each of 12 plates, and the nick-translated 438 bp fragment of the rat cDNA for β-polymerase was used as probe (5). Twenty-eight common signals eventually were identified in the autoradiograms, and areas of the plates corresponding to these positive phage were removed and resuspended in SM buffer (50 mM Tris-HCl, pH7.5, 0.1M NaCl, 10 mM $MgCl_2$, and 0.01% gelatin). The phage were replated at low plaque density and then the plates were screened as before. Positive phage from the second screening were isolated and plaque purified. Finally, phage DNA was prepared as described by Cobianchi et al (13).

Rabbit anti-chicken DNA polymerase β IgG was used for testing the plaque purified clone, λpolβ-h2, at a concentration of 20 µg/ml. This solution was capable of detecting 0.1 ng of purified mammalian DNA polymerase β spotted directly on a filter. Goat anti-rabbit IgG horseradish peroxidase conjugate (Bio-Rad; diluted 1:200) was used (14) for identification of positive clones.

Subcloning of β-polymerase cDNA fragments and DNA sequencing. The cDNA insert from clone λpolβ-h2 was isolated by cleavage with restriction endonuclease EcoRI followed by electrophoresis in a low melting agarose gel. The two fragments, faster and slower migrating, respectively, were ligated with pUC9 that had been EcoRI cleaved and phosphatase treated. The resulting recombinants, termed pUC9-2S and pUC9-2F, then were used to transform E. coli JM-83. cDNAs from pUC9-2S and pUC9-2F were subcloned in M13 and sequenced according to the method described by Sanger et al. (15).

RNA blot hybridization. PolyA$^+$ RNA of a human teratocarcinoma cells, NTera2D1, in growth phase was prepared as described (11). PolyA$^+$ RNA was denatured by heating at 60°C for 15 min in 2.2 M formaldehyde, 50% (vol/vol) formamide and then subjected to electrophoresis in a 1.5% agarose gel containing 2.2 M formaldehyde, 20 mM morpholinopropane sulfonate, pH 7.0, 5 mM sodium acetate, 1 mM EDTA. E. coli and mammalian rRNAs were used as molecular weight markers. After electrophoresis gels were rinsed with water, soaked in 15 mM NaOH, 10 mM NaCl for 15 min at room temperature, neutralized in 0.1 M Tris-HCl, pH 7.5, for 45 min, then soaked in 0.025 M Na_2HPO_4, pH 6.5, for 1 h. RNA was electrotransferred to Gene-Screen Hybridization Transfer Membrane and hybridized to nick-translated probe for 10-24 h at 42° in 50% formamide, 0.02% polyvinylpyrrolidone (M_r 40,000), 5xSSC, 1% SDS and denatured salmon sperm DNA (0.3 mg/ml). Membranes were exposed to Kodak XAR-5 film.

Results

Isolation of human β-polymerase cDNA clones

The β-polymerases from mammalian sources have related structures (16,17). Thus, we assumed that a cDNA for rat β-polymerase could be used as a probe for selecting β-polymerase cDNAs from a human library. Our approach was to first screen a human cDNA library in λgtll using a fragment of the rat cDNA and then to test positive clones for production of a polypeptide capable immunobinding with a β-polymerase specific antibody. PolyA$^+$ RNA from a human teratocarcinoma cell line was used to construct a cDNA library in the EcoRI site of phage λgtll, and the library was screened with a 438 bp EcoRI fragment of the rat β-polymerase cDNA, λpolβ-10 (5); we had found earlier that this probe hybridized with polyA$^+$ RNA from the same cells. A detailed characterization of the cDNA for rat β-polymerase is described elsewhere (5). Twenty-eight positive phage were identified after screening a total of 2.4×10^7 recombinant phage. Five of them eventually were obtained as purified clones, and one clone, designated λpolβ-h2, was selected for detailed study. This clone was found to be positive when tested with the polyclonal antibody to β-polymerase that was used earlier for isolation of the rat β-polymerase cDNA.

Phage λpolβ-h2 contained a 1257 bp insert with one internal EcoRI site leading to fragments of 564 and 699 bp, respectively. A partial restriction map and strategy for sequencing of the insert are shown in Figure 1. The two RI fragments were subcloned into pUC9 and then into M13. Sequencing was conducted using the method described by Sanger et al. (15). Two or more sequencing determinations were conducted with all regions of the insert.

Primary Structure of human β-polymerase cDNA

The nucleotide sequence of the cDNA insert in λpolβ-h2 is shown in Figure 2, along with the sequence of the protein deduced from the long open reading frame. Nucleotide sequence and codon differences between the human and rat

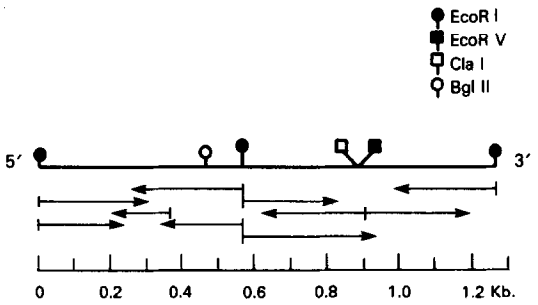

Figure 1. Diagram illustrating partial restriction map and sequencing strategy for the insert in λpol β-h2. Sequencing by the method described by Sanger (15) was performed with synthetic oligonucleotide primers based upon the sequence of the rat β-polymerase cDNA. Alternatively, sequencing by the same method was conducted after subcloning the two EcoRI fragments in M13 and use of the usual M13-complementary oligonucleotide primers.

```
      ccggagctgggttgctcctgctcccgtctccaagtcctggtacctccttcaagctgggagagggctcta   69

gtccctggttctgaacactctgggggttctcgggtgcaggccgccatgagcaaacggaaggcgccgcagagactct  144

                            Val
                            GTG                                          G
caacgggggaatcaccgac ATG CTC ACA GAA ATA GCA AAC TTT GAG AAG AAC GTG AGC CAA  205
                    Met Leu Thr Glu Leu Ala Asn Phe Glu Lys Asn Val Ser Gln   14

  G                         A                           G           G
GCT ATC CAC AAG TAC AAT GCT TAC AGA AAA GCA GCA TCT GTT ATA GCA AAA TAC CCA  262
Ala Ile His Lys Tyr Asn Ala Tyr Arg Lys Ala Ala Ser Val Ile Ala Lys Tyr Pro   33

          C               A                           A
CAC AAA ATA AAG AGT GGA GCT GAA GCT AAG AAA TTG CCT GGA GTA GGA ACA AAA ATT  319
His Lys Ile Lys Ser Gly Ala Glu Ala Lys Lys Leu Pro Gly Val Gly Thr Lys Ile   52

                        A
GCT GAA AAG ATT GAT GAG TTT TTA GCA ACT GGA AAA TTA CGT AAA CTG GAA AAG ATT  376
Ala Glu Lys Ile Asp Glu Phe Leu Ala Thr Gly Lys Leu Arg Lys Leu Glu Lys Ile   71

                                                        Thr
  T               A                       C                       C   G
CGG CAG GAT GAT ACG AGT TCA TCC ATC AAT TTC CTG ACT CGA GTT AGT GGC ATT GGT  433
Arg Gln Asp Asp Thr Ser Ser Ser Ile Asn Phe Leu Thr Arg Val Ser Gly Ile Gly   90

                        Leu
                        C                                               G
CCA TCT GCT GCA AGG AAG TTT GTA GAT GAA GGA ATT AAA ACA CTA GAA GAT CTC AGA  490
Pro Ser Ala Ala Arg Lys Phe Val Asp Glu Gly Ile Lys Thr Leu Glu Asp Leu Arg  109

                                                                Glu
                                C                               A
AAA AAT GAA GAT AAA TTG AAC CAT CAT CAG CGA ATT GGG CTG AAA TAT TTT GGG GAC  547
Lys Asn Glu Asp Lys Leu Asn His His Gln Arg Ile Gly Leu Lys Tyr Phe Gly Asp  128

          G                     G         C G                     A
TTT GAA AAA AGA ATT CCT CGT GAA GAG ATG TTA CAA ATG CAA GAT ATT GTT CTA AAT  604
Phe Glu Lys Arg Ile Pro Arg Glu Glu Met Leu Gln Met Gln Asp Ile Val Leu Asn  147

                Leu     Pro
              G C       C     G       C                               C
GAA GTT AAA AAA GTG GAT TCT GAA TAC ATT GCT ACA GTC TGT GGC AGT TTC AGA AGA  661
Glu Val Lys Lys Val Asp Ser Glu Tyr Ile Ala Thr Val Cys Gly Ser Phe Arg Arg  166

                                                        Asn
  C                     A   T       C       G           C   A  A
GGT GCA GAG TCC AGT GGT GAC ATG GAT GTT CTC CTG ACC CAT CCC AGC TTC ACT TCA  718
Gly Ala Glu Ser Ser Gly Asp Met Asp Val Leu Leu Thr His Pro Ser Phe Thr Ser  185

        Ser                             Arg
          G                 G             GT          A               A   C
GAA TCA ACC AAA CAG CCA AAA CTG TTA CAT CAG GTT GTG GAG CAG TTA CAA AAG GTT  775
Glu Ser Thr Lys Gln Pro Lys Leu Leu His Gln Val Val Glu Gln Leu Gln Lys Val  204

Arg
 G   C                T
CAT TTT ATC ACA GAT ACC CGT TCA AAG GGT GAG ACA AAG TTC ATG GGT GTT TGC CAG  832
His Phe Ile Thr Asp Thr Leu Ser Lys Gly Glu Thr Lys Phe Met Gly Val Cys Gln  223

           Glu             Asn
         C G G               C                   G       C               G
CTT CCC AGT AAA AAT GAT GAA AAA GAA TAT CCA CAC AGA AGA ATT GAT ATC AGG TTA  889
Leu Pro Ser Lys Asn Asp Glu Lys Glu Tyr Pro His Arg Arg Ile Asp Ile Arg Leu  242

Ile
AT    C                 C                       C           A       C   C   T
TAC CCA AAA GAT CAG TAT TAC TGT GGT GTT CTC TAT TTC ACT GGG AGT GAT ATT TTC  946
Tyr Pro His Asp Gln Tyr Tyr Cys Gly Val Leu Tyr Phe Thr Gly Ser Asp Ile Phe  261

                                Leu
                  G   G         CT              C
AAT AAG AAT ATG AGA GCT CAT GCC AAG GAA AAG GGT TTC ACA ATC AAT GAG TAC ACC 1003
Asn Lys Asn Met Arg Ala His Ala Lys Glu Lys Gly Phe Thr Ile Asn Glu Tyr Thr  280

      C     C     G           G       T   G   G       T   G       C   C   G
ATC CGT CCC TTG GGA GTC ACT GGA GTT GCA GGA GAA CCC CTG CCA GTG GAT AGT GAA 1060
Ile Arg Pro Leu Gly Val Thr Gly Val Ala Gly Glu Pro Leu Pro Val Asp Ser Glu  299

Gln                             Arg
CAG       T   C                 CGC
AAA GAC ATC TTT GAT TAC ATC CAG TGG AAA TAC CGG GAA CCC AAG GAC CGG AGC GAA 1117
Lys Asp Ile Phe Asp Tyr Ile Gln Trp Lys Tyr Arg Glu Pro Lys Asp Arg Ser Glu  318

tgaggcctgtatcctccctggcgcagacacaacccaatgggtcttaatttatttcttaacctttgctatgtaagg 1192
***

gtctttggtgtttttaaatgattgtttcttcttcatgcttttgcttgcaatgtagtcaataaaacc          1258
```

Figure 2. The nucleotide sequence of the insert in λpolβ-h2; the sequence of the protein encoded by the long open reading frame, and comparison with the coding sequence of the rat β-polymerase cDNA (5). Nucleotides are numbered starting from the 5' end of the human cDNA and extending to the 3' end residue, 1257. Differences between the human and rat cDNAs are indicated on the upper two lines of each row by the nucleotide and amino acid of the rat cDNA. The human sequence extends 122 residues 5' of the start (|) of the rat cDNA.

cDNAs are indicated. The two cDNAs were similar, except that the human sequence extended 122 residues 5' of the end of the rat cDNA. The reading frame and translation start codon of the human mRNA were assigned by comparison with the rat β-polymerase cDNA. The start codon was the first in-frame ATG from the 5' end corresponding to residues 164, 165 and 166. The 36 residue sequence preceding this codon contained 18S rRNA complementary sequences, as seen in several mammalian mRNAs (18). Overall the 5' untranslated region of the human cDNA contained 163 bp. This sequence had one ATG triplet, but this was not in-frame with the coding region. There was onein-frame termination codon between the 5' end of the cDNA and the first in-frame ATG codon (residues 164-166). The 3' untranslated region of the human cDNA did not contain a polyA sequence, but did have a mRNA processing signal (AATAAA) starting 9 bases from the end. This putative mRNA processing signal corresponded in location, relative to the termination codon, with the processing signal in the rat cDNA. Over the 144 base sequence of the 3' untranslated region of the human cDNA, 64% of the residues matched with residues in the 3' untranslated region of the rat cDNA. By contrast, the other portion of the human cDNA matched in 90% of residues with the rat cDNA.

The human cDNA contained one long open reading frame that was the same length as that of the rat β-polymerase cDNA (954 bases). The human protein deduced from the long open reading frame contained 318 amino acids (M_r=36,133), 302 of which matched with residues in the deduced rat protein. The other two reading frames of the human cDNA contained 20 and 30 termination codons, respectively. In the long open reading frame, the human and rat cDNAs differed at 92 residues and 70 of these differences did not lead to a codon difference. Two unusual amino acid differences were noted, Ileu:Tyr and Leu:Lys in rat and human, respectively.

Northern blot analysis of human teratocarcinoma cell RNA

A probe prepared with the 5' EcoRI fragment of the β-polymerase cDNA was hybridized with a Northern blot of polyA$^+$ RNA from growing NTera2D1 cells. A single prominent band of hybridizing RNA was detected corresponding to a molecule of approximately 1400 bases (Figure 3). The size of this RNA species was close to the size of the λpolβ-h2 cDNA itself, and a species of ~1400 bases was expected, assuming the presence of a 100 base polyA sequence for the mRNA copied.

Discussion

The cloned phage λpolβ-h2 produced a peptide that was reactive with a β-polymerase specific antibody, and the sequence of the insert (1257 bp) closely matched the sequence of a previously characterized rat cDNA for β-polymerase (1184 bp). The identity of the rat cDNA had been confirmed by a

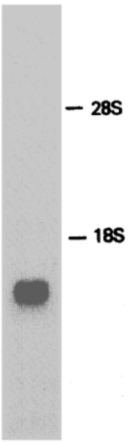

Figure 3. Autoradiogram showing results of Northern blot of polyA$^+$ RNA of growing NTera2D1 cells. 10 µg RNA was electrophoresed in an agarose gel after denaturation in a solution containing 50% formamide, 2.2 M formaldehyde, as described (13). The probe was the 5' EcoRI fragment of the λpolβ-10 insert ^{32}P labeled by nick translation to a specific activity of 1 x 10^8 dpm/µg. Exposure was for 48h.

match of the partial sequence of the protein (15% of total primary structure) with the sequence of the cDNA (5), and this same primary structure match is evident from the sequence of the protein deduced from the cDNA in λpolB-h2. It is clear, therefore, that the λpolβ-h2 insert is a cDNA for human β-polymerase.

The cloning and sequencing of a human β-polymerase cDNA was undertaken to provide probes and sequence information necessary to study the structure and function of the human β-polymerase gene. The library from teratocarcinoma cells was chosen because it was constructed to contain a high concentration of full-length cDNA. The insert in λpolβ-h2 contained an open reading frame analogous to the coding region of the rat β-polymerase cDNA, as well as 163 residues 5' of the ATG start codon. These residues contained one in-frame termination codon, but no in-frame start codons. We conclude from this sequence information, the match with the rat cDNA, and the results of the Northern blot analysis (Figure 5), that the human cDNA described here contains the entire coding region for human β-polymerase plus most of the 5' and 3' untranslated regions.

There is a high degree of similarity between the cDNA deduced sequences of rat and human β-polymerases. Both proteins contained 318 amino acids and 302 were identical. Most of the nucleotide differences in the coding region of the two cDNAs did not lead to an amino acid codon difference, and in the 16 cases where a codon difference occurred, 9 lead to an amino acid of similar chemical properties. The number of nucleotide differences (92) between the 954 residue coding regions permits a calculation of the substitution rate, assuming the

sequences diverged 80×10^6 years ago (19). An overall substitution rate of 1.2 per site per 10^9 year was obtained. This value is quite low and is about one-half that of other mammalian coding regions characteristic of low overall substitution rates (HPRT, ACTH, Histone H2A, and parathyroid hormone) (19). The ratio of nonconservative to overall nucleotide differences(0.24) suggests that β-polymerase is subject to fairly strong selective pressure, though not as strong as that for α-actin, histones, or helix-destabilizing protein (13,19). Plots of computer-derived secondary structure predictions for human β-polymerase were similar to those for rat β-polymerase (5) and indicated 7 regions of α-helix spaced throughout the sequence and several regions of β-sheet (not shown).

Since the insert described here appears to be full-length for the coding region, it should be possible to produce large amounts of active native enzyme in E. coli through use of appropriate expression vector systems. This would greatly facilitate detailed structure-function studies of β-polymerase, a research area of particular promise since this enzyme is the simplest naturally occurring DNA polymerase known (20).

References

1. Weissbach, A. (1979) Arch. Biochem. Biophys. 198, 386-396.
2. Fry, M. (1982) in Enzymes of Nucleic Acid Synthesis and Modification, (Jacob, S.T., Ed.) Vol. 1, CRC Press, Boca Raton, FL.
3. Miller, M.R., and Chinault, D.W. (1982) J. Biol. Chem. 257, 10204-10209.
4. Paterson, M.C., and Smith, P.J. (1979) Annu. Rev. Genet. 13, 291-318.
5. Zmudzka, B.Z., SenGupta, D.N., Matsukage, A., Cobianchi, F., Kumar, P., and Wilson, S.H. (1986) Proc. Natl. Acad. Sci., USA., in press.
6. McBride, O.W., Zmudzka, B.Z., and Wilson, S.H. (1986), submitted for publication.
7. Andrews, P.W., Damjanov, I., Simon, D., Banting, G., Dracopoli, N.C. and Fogh, J. (1984) Lab. Invest. 50, 147-162.
8. Gerard, G.F. (1985) Focus 7, 1-3.
9. Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161-170.
10. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) DNA Cloning (Glover, D.M., Ed.) Vol. 1, IRL Press, Oxford, Washington, DC, pp. 49-78.
11. Skowronski, J., and Singer, M.F. (1985) Proc. Natl. Acad. Sci. USA 82, 6050-6054.
12. Benton, W.D., and Davis, R.W. (1977) Science 196, 180-182.
13. Cobianchi, F., SenGupta, D.N., Zmudzka, B.Z., and Wilson, S.H. (1986) J. Biol. Chem., 261, In Press.
14. Swack, J.A., Karawya, E., Albert, W., Fedorko, J., Minna, J.D., and Wilson, S.H. (1985) Anal. Biochem. 147, 10-21.
15. Sanger, F. (1980) J. Mol. Biol. 143, 161-178.
16. Chang, L.M.S., Plevani, P., and Bollum, F.J. (1982) Proc. Natl. Acad. Sci. USA 79, 758-761.
17. Tanabe, K., Yamaguchi, T., Saneyoshi, M., Matsukage, A., and Takahashi, T. (1984) J. Biochem. 96, 365-370.
18. Lewin, B. (1985) Genes, John Wiley and Sons, New York, NY.
19. Li, W.-H., Wu, C.-I., and Lao, C.-C. (1985) Mol. Biol. Evol. 2, 150-174.
20. Tanabe, K., Bohn, E.W., and Wilson, S.H. (1979) Biochemistry 18, 3401-3406.