MOLECULAR AND CELLULAR BIOLOGY, Dec. 1992, p. 5673–5682
0270-7306/92/125673-10$02.00/0

Vol. 12, No. 12

# Sequence of cDNA Comprising the Human *pur* Gene and Sequence-Specific Single-Stranded-DNA-Binding Properties of the Encoded Protein

ANDREW D. BERGEMANN, ZHI-WEI MA, AND EDWARD M. JOHNSON*
*Department of Pathology, Mount Sinai Medical School, New York, New York 10029*

Received 30 June 1992/Returned for modification 12 August 1992/Accepted 17 September 1992

**The human Pur factor binds strongly to a sequence element repeated within zones of initiation of DNA replication in several eukaryotic cells. The protein binds preferentially to the purine-rich single strand of this element, *PUR*. We report here the cloning and sequencing of a cDNA encoding a protein with strong affinity for the *PUR* element. Analysis with a series of mutated oligonucleotides defines a minimal single-stranded DNA Pur-binding element. The expressed Pur open reading frame encodes a protein of 322 amino acids. This protein, Purα, contains three repeats of a consensus motif of 23 amino acids and two repeats of a second consensus motif of 26 amino acids. Near its carboxy terminus, the protein possesses an amphipathic α-helix and a glutamine-rich domain. The repeat region of Pur cDNA is homologous to multiple mRNA species in each of several human cell lines and tissues. The HeLa cDNA library also includes a clone encoding a related gene, Purβ, containing a version of the 23-amino-acid consensus motif similar, but not identical, to those in Purα. Results indicate a novel type of modular protein with capacity to bind repeated elements in single-stranded DNA.**

The Pur factor has recently been identified as a HeLa cell nuclear protein which binds a sequence element adjacent to a region of stably-bent DNA upstream of the human c-*myc* gene (6). This element is near the center of a region implicated as an initiation zone for chromosomal DNA replication (36, 56). Pur binds preferentially to the purine-rich single-stranded DNA form of its recognition element. Binding to single-stranded oligonucleotides is highly specific and is markedly reduced when the binding sequence is altered. The Pur-binding element is conserved in gene-flanking regions and origins of DNA replication throughout eukaryotes. A version of the *PUR* element adjacent to a bend in a reported hamster *dhfr* replication initiation zone (13) effectively competes with the c-*myc* version for binding to the HeLa cell nuclear factor (6).

Several proteins from prokaryotes have previously been characterized which play a role in DNA replication and which bind to single-stranded DNA. The amino acids involved in DNA binding by certain of these proteins have been identified (24). However, none of these single-stranded-DNA-binding proteins is known to have any sequence specificity. In eukaryotes, the single-stranded-DNA-binding protein RP-A is required for replication initiated at the simian virus 40 origin in vitro (18, 59). No sequence specificity has been reported for DNA binding by RP-A. Recently, Hofmann and Gasser (30) have reported the purification of a protein, ACBP, which binds the yeast autonomously replicating sequence (ARS) consensus sequence. ACBP binds specifically to the T-rich single-stranded form of the ARS consensus sequence, and binding to the histone H4 ARS correlates with H4 ARS activity. These results, together with observations on Pur, raise the possibility that sequence-specific single-stranded-DNA-binding proteins may serve an important function in the initiation of DNA replication in eukaryotic cells. At this time, however, no amino acid sequence has been reported for any eukaryotic protein known to bind specifically to a single-stranded DNA element.

We report here the cloning of a human fetal liver cell cDNA which encodes a protein that specifically binds the purine-rich single strand of the *PUR* element located upstream of the c-*myc* gene. Northern (RNA) analysis reveals that multiple transcripts homologous to this cDNA are present in a variety of human cell types. Analysis of the expressed open reading frame of Pur reveals a modular structure unique among known DNA-binding proteins.

## MATERIALS AND METHODS

**Oligonucleotides used.** Oligonucleotide MF0677 (GGAG GTGGTGGAGGGAGAGAAAAG) is a 24-mer representing the c-*myc* sequence element initially established as binding to Pur (6). The oligonucleotides used as competitors with MF0677 in binding experiments are described in Table 1. The following two oligonucleotides were used as nonspecific competitors in the screening of expression libraries: MR0740, TCTCAAGCTTGGTCCCTCAC, and MF0562, TACTGAATTCACTTAACACT. Oligonucleotides used as primers in 3′ and 5′ rapid amplification of cDNA ends (RACE) experiments are listed in the legend to Fig. 7.

**Screening of expression libraries for proteins with affinity for the labeled, single-stranded *PUR* element.** A human fetal liver cDNA library (Clontech catalog no. HL1005, prepared in vector λgt11) was screened for clones which bind the purine-rich sequence from upstream of the c-*myc* gene (oligonucleotide MF0677), essentially as described previously for double-stranded oligonucleotides (51, 52). The library was plated on *Escherichia coli* Y1090 on six 150-mm plates at a density of $5 \times 10^4$ PFU per plate. Plates were incubated at 42°C for 3 h. The plates were then overlaid with nitrocellulose filters which had previously been saturated with 10 mM IPTG (isopropyl-β-D-thiogalactopyranoside) and then dried. The plates were incubated for a further 6 h at

* Corresponding author.

TABLE 1. Binding of various synthetic oligonucleotides to protein extract of λAB6 lysogen[a]

| Competitor[b] | Sequence | % Binding activity remaining[c] | |
|---|---|---|---|
| | | 5-fold excess | 20-fold excess |
| MF0677 | GGAGGTGGTGGAGGGAGAGAAAAG | 45 | 16 |
| MC0677 | GGAGGTGGTGGAGGGTTTTTTTTT | 61 | 10 |
| ME0677 | GGAGGTGGTGGAGGTTTTTTTTTT | 67 | 44 |
| MG0677 | GGAGGTGGTGGAGTTTTTTTTTTT | 100 | 76 |
| MH0677 | TTTTTTTTTGGAGGGTTTTTTTTT | 82 | 48 |
| MI0677 | TTTTTTGGTGGAGGGTTTTTTTTT | 76 | 35 |
| MJ0677 | TTTGGTGGTGGAGGGTTTTTTTTT | 55 | 15 |
| DR3529 | TGATGAGGGAGAGGGAGAAGGGAT | 85 | 74 |
| Poly(G) | GGGGGGGGGGGGGGGGGGGGGGGG | 93 | 85 |
| Poly(A) | AAAAAAAAAAAAAAAAAAAAAAAA | 97 | 99 |

[a] Binding reactions were performed as gel shift assays as described in Materials and Methods. The labeled probe in each reaction was oligonucleotide MF0677. Competing oligonucleotides were added in either 5-fold or 20-fold excess.

[b] Single-stranded oligonucleotides used were the following: MF0677, the 24-mer representing the *PUR* element site originally detected upstream of the human c-*myc* gene; MC0677, ME0677, MG0677, MH0677, MI0677, and MJ0677, mutants of oligonucleotide MF0677 in which indicated residues have been replaced by T; DR3529, a 24-mer representing the hamster *dhfr* version of the *PUR* element; Poly(G) and Poly(A), homopolymers representing the two most common bases in the *PUR* element.

[c] After autoradiography of the gel shift assay results, bands were scanned with a Xerox Data Copy GS+ densitometer.

37°C. After being lifted from the plates, filters were immersed in binding buffer (50 mM KCl, 10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES] [pH 7.8], 0.2 mM EDTA, 0.5 mM dithiothreitol) containing 5% nonfat milk powder (Carnation) and gently shaken at room temperature for 60 min. Filters were then washed for 5 min at room temperature three times in binding buffer. Filters were then shaken gently for 60 min at room temperature in 15 ml of binding buffer containing $10^6$ cpm of probe (MF0677 labeled by polynucleotide kinase) and 5 μg each of MF0562 and MR0740 per ml. Filters were washed four times in binding buffer (50 ml per filter) at room temperature for 7 min and then blotted dry before autoradiography. Potential positives were pooled and rescreened.

**Determination and analysis of the nucleotide sequence of Purα.** Inserts in λgt11 were cloned into M13 bacteriophage and sequenced by the dideoxy termination technique (50). Sequence analysis was conducted by using the IBI Pustell programs. The sequence of Purα obtained from clones AB6 and HE1 has been registered with the GenBank sequence data bank under accession number M96684, dated 8 July 1992.

**DNA binding studies of Pur-LacZ fusion protein.** Y1090 clones lysogenized with λAB5 and λAB6 were isolated by infection at a multiplicity of infection of 10 and then plated on agar containing Luria broth and incubated overnight at 32°C as previously described (51). Lysogens were identified by their inability to grow at 42°C. To prepare protein extracts, we grew lysogens at room temperature with shaking in Luria broth until they reached the mid-log phase. IPTG was then added to 5 mM, and the cultures were incubated at 37°C for 2 h with shaking. Aliquots of 1.0 ml of each culture were pelleted at 13,000 × *g* for 1 min, and each pellet was resuspended in 50 μl of extraction buffer (10 mM HEPES [pH 7.9], 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride). Each sample was sonicated six times for 5 s at setting 4 on a VibraCell sonicator with a microtip probe. To each such protein extract, 40 μl of 50% glycerol plus 10 μl of 5 M NaCl were added, and extracts were then incubated for 15 min at 4°C. The extracts were cleared by a 30-min centrifugation at 13,000 × *g* at 4°C. Protein extracts were prepared from phage lysates as follows. Phage were plated on 150-mm plates of Y1090 at a density of $10^5$ per plate and incubated at 37°C for 9 h. The top agar was collected, mixed with 10 ml of extraction buffer, and incubated at 4°C for 1 h. The agar was then removed by centrifugation at 12,000 × *g* for 10 min, and the supernatant was dialyzed against extraction buffer for 6 h at 4°C. Aliquots (2 μl) of each protein extract were used in gel shift assays as previously described for nuclear extracts (6).

**Tissue culture, poly(A)$^+$ RNA preparation, and hybridization analysis.** Cell lines used for the preparation of mRNA were grown in suspension in 1-liter Corning Spinner flasks. HeLa cells were grown in Dulbecco modified Eagle medium (GIBCO), human hepatoma cells (HepG2) were grown in Joklik modified Eagle medium (Sigma), and small-cell lung carcinoma cells (NCI-N82) were grown in RPMI 1640 medium (GIBCO). All media were supplemented with 10% fetal calf serum, 100 U of penicillin G per ml, and 100 μg of streptomycin per ml. Cells were pelleted at 150 × *g* for 5 min. Human fetal liver tissue, second trimester from spontaneous abortuses, was kindly supplied by Sunkara Rao, Renata Dische, and Stave Kohtz. For preparation of RNA, tissue was frozen in liquid nitrogen and ground in a Biopulverizer. Total RNA was extracted from both tissues and harvested cells as previously outlined (10). RNA samples were passed twice through oligo(dT)-cellulose columns to purify poly(A)$^+$ RNA (4). The RNA concentration of each sample was determined by the $A_{260}$. Poly(A)$^+$ RNA (10 μg per lane) was subjected to electrophoresis on 1.2% agarose gels containing 2.2% formaldehyde at 2.5 V/cm for 3.75 h (49). RNA was transferred to GeneScreen Plus membranes (Du Pont), exposed to UV light for 2 min, and baked at 80°C in a vacuum oven for 2 h. Filters were prehybridized for 3 h at 70.5°C in Northern hybridization solution [50 mM *N*-tris(hydroxyl)methyl-2-aminoethanesulfonic acid [TES] [pH 7.0], 25 mM $Na_2HPO_4$, 25 mM $NaH_2PO_4$, 0.3 M NaCl, 30 mM trisodium citrate, 1% sodium dodecyl sulfate [SDS], 10% dextran sulfate, 100 μg of sonicated salmon sperm DNA per ml, 100 μg of yeast RNA per ml), and hybridization to the probe was performed in the same solution at 70.5°C for 17 h. The probe used was the 777-bp *Pst*I fragment of the Purα cDNA (nucleotides 165 through 941), labeled with $^{32}$P-phosphate by a random priming procedure (19). After hybridization, filters were washed twice for 5 min each in 2× SSC–1% SDS (SSC is 150 mM NaCl plus 15 mM sodium citrate, pH 7.0) at 70.5°C and once for 30 min in 0.1× SSC–1% SDS at 70.5°C. Filters were then rinsed in 2× SSC and autoradiographed on Kodak XAR film.

**5′ and 3′ extension of Purα mRNA.** HepG2 cell poly(A)$^+$ RNA prepared as described above was used for RACE. Rapid amplification of the 3′ cDNA end of *pur* was done essentially as described previously (21). Positions of the primers used are detailed in the legend to Fig. 7. HepG2 cell poly(A)$^+$ RNA, 1.0 μg in 13 μl of distilled water, was denatured at 70°C for 5 min and immediately chilled on ice. The first-strand cDNA was synthesized with 1.0 μM primer PDT-01 and 200 U of reverse transcriptase from the Super-Script Preamplification System (Bethesda Research Laboratories) according to the manufacturer's instructions, except that 20 U of RNasin (Promega) was included in the reaction mix. cDNA (1 μl) from the reverse transcription reaction

mix was then polymerase chain reaction (PCR) amplified in a 100-μl reaction volume with the addition of 2.5 U of *Taq* polymerase, using two primers (PDT-01 and EX-695, each 1.0 μM). The PCR profile was as follows: denaturation at 94°C for 1 min, annealing reaction at 54°C for 2 min, and extension at 72°C for 5 min. A 1-μl sample of the first amplification reaction was used for the second amplification of the 3′ cDNA end of *pur*. Primer EX-990 (1.0 μM) was substituted for primer EX-695. The PCR was performed for 20 cycles in the same reaction buffer and volume and with the same profile as described above. Rapid amplification of the 5′ cDNA end of *pur* was done as follows. HepG2 poly(A)$^+$ RNA (1 μg) was reverse transcribed as described above, except for substitution of 1.0 μM primer EX-270 for PDT-01. The cDNA was purified from excess primer by passing it through a Centricon 100 microconcentrator (Amicon). The first-strand cDNA synthesized above was polyadenylated at the 3′ end by terminal deoxynucleotidyl transferase (Bethesda Research Laboratories) in the presence of 100 μM dATP at 37°C for 30 min. The cDNA was incubated at 70°C for 2 min, precipitated with ethanol, and redissolved in 20 μl of TE (pH 7.5). A 1-μl sample of the reaction mix thus generated was used for amplification with 2.5 U of *Pyrococcus* DNA polymerase (Stratagene) with primers EX-270 and PDT-01 (each 1.0 μM) in 100 μl as described above. The PCR profile was as follows: denaturation at 97.5°C for 1 min, annealing reaction at 54°C for 2 min, extension at 74°C for 3 min. After 15 cycles, an additional 2.5 U of *Pyrococcus* polymerase was added, and the reaction was continued for an additional 15 cycles. A 1-μl sample of the first amplification reaction mix thus obtained was used for the second amplification reaction as described above except that primer EX-174 was substituted for EX-270. Samples of 10 μl of each of the RACE products were separated on a 1.5% agarose gel and transferred to a GeneScreen Plus membrane. Membranes were exposed to UV light for 2 min and prehybridized in the DNA prehybridization solution (same as above, but with 3× SSC and without yeast RNA) at 65°C for 2 h. The probe was the $^{32}$P-phosphate-labeled *pur* insert (1,080 bp) of λAB6. Hybridization was performed as described above at 65°C for 6 h. The membranes were then washed twice in 2× SSC–1% SDS at 65°C each for 5 min, twice in 0.1× SSC–1% SDS at 65°C each for 30 min, and once for 5 min in 0.1× SSC and autoradiographed as described above.

**Screening libraries by DNA-DNA hybridization.** A λZAP II HeLa cell cDNA library was kindly supplied by J. L. Manley. The library was screened for clones of Pur by previously described techniques (5, 44, 54). The 777-bp *Pst*I fragment, outlined for use in Northern analysis, was labeled by a random priming technique (19) and hybridized to the filters in hybridization solution (20 mM Tris-HCl, 10% dextran sulfate, 24 mM sodium phosphate, 0.1% SDS, 750 mM NaCl, 75 mM trisodium citrate, pH 7.6) at 60°C for 16 h. Filters were rinsed five times in 2× SSC at room temperature, washed in 2× SSC for 15 min at 60°C, and then washed for 10 min at 60°C in wash buffer (40 mM sodium phosphate buffer [pH 7.2], 1 mM EDTA, 5% SDS). Finally, the filters were rinsed twice in 2× SSC, dried, and autoradiographed.

## RESULTS

**Isolation of Pur clones from a λgt11 expression library based on affinity for the *PUR* element.** Screening of 3 × 10$^5$ plaques from a human fetal liver cDNA library in λgt11 with labeled MF0677 oligonucleotide, which represents the purine-rich strand of the c-*myc PUR* element, as described in Materials and Methods, yielded several positive plaques which were pooled and rescreened. Six individual plaques (designated λAB1 through λAB6) which were positives in the second screening were each replated onto separate plates and lifted onto new filters. Each filter was then cut in half. The separate halves were screened for binding of the labeled oligonucleotide MF0677, either in the presence of 10 μg of unlabeled MF0677 per ml (Fig. 1, right) or in the presence of 10 μg of an unlabeled nonspecific competitor oligonucleotide, MF0562, per ml (Fig. 1, left). For all six clones, a result was obtained equivalent to that displayed for λAB2 in Fig. 1. It can be seen that excess MF0677 specifically competes for binding but excess MF0562 does not, indicating the sequence-specific nature of binding to the single-stranded *PUR* element. All six clones contained the same 1.08-kb insert.

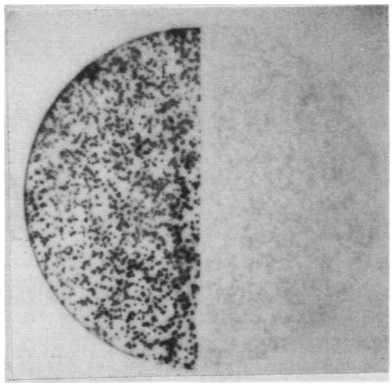

FIG. 1. Specific binding of the *PUR* element by protein in plaques of phage λAB2. λAB2 phage were plated, induced with IPTG as described in Materials and Methods, and transferred to nitrocellulose membranes. Each membrane was then cut in half, and the halves were incubated separately in the presence of MF0677 oligonucleotide, which was $^{32}$P end labeled as described in Materials and Methods. The left half was incubated in the presence of a 20-fold excess of unlabeled nonspecific competitor MF0562, while the right half was incubated in the presence of a 20-fold excess of unlabeled specific competitor MF0677.

**Sequence analysis of Pur.** The nucleotide sequence of the *pur* gene insert in λAB6 was determined after subcloning from λgt11 into M13 as described in Materials and Methods. The sequence of the cDNA is presented in Fig. 2 together with the deduced amino acid sequence of the Pur protein. The orientation of the insert in λAB6 was determined by PCR analysis, and the open reading frame displayed beneath the sequence in Fig. 2 forms a fusion protein with the β-galactosidase gene of the λgt11 vector. The open reading frame, designated Purα, is 322 amino acids in length. Codon 4 is an ATG, coding for a methionine, that nearest the 5′ end of the insert. The sequencing of another cDNA clone, λHE1, extends the open reading frame an additional 20 codons, none of which are methionines. The sequence derived from both clones is shown in Fig. 2. The 5′ terminus of clone λAB6 is indicated by a vertical line below the sequence shown. For several reasons, we believe the methionine codon denoted position 1 in λAB6 represents the human translation start site for the Pur protein. First, the protein encoded by λAB6 possesses full single-stranded-DNA-binding capacity with a specificity similar to that of the Pur protein observed in HeLa cells. Second, codon usage is strongly representative of that preferred in human cells throughout the open reading frame but is significantly less

```
-59 CGACTGAGGCGGCGGGCGGAGCGGCAGGCGGCGGCGGCGCGGCAGCGGAGCGCAGCATC

  1 ATG GCG GAC CGA GAC AGC GGC AGC GAG CAG GGT GGT GCG GCG CTG
    Met Ala Asp Arg Asp Ser Gly Ser Glu Gln Gly Gly Ala Ala Leu

 46 GGT TCG GGC GGC TCC CTG GGG CAC CCC GGC TCG GGC TCA GGC TCC
    Gly Ser Gly Gly Ser Leu Gly His Pro Gly Ser Gly Ser Gly Ser

 91 GGC GGG GGC GGT GGT GGC GGC GGG GGC GGC GGC GGC AGT GGC GGC
    Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Ser Gly Gly

136 GGC GGC GGC GGG GCC CCA GGG GGG CTG CAG CAC GAG ACG CAG GAG
    Gly Gly Gly Gly Ala Pro Gly Gly Leu Gln His Glu Thr Gln Glu

181 CTG GCC TCC AAG CGG GTG GAC ATC CAG AAC AAG CGC TTC TAC CTG
    Leu Ala Ser Lys Arg Val Asp Ile Gln Asn Lys Arg Phe Tyr Leu

226 GAC GTG AAG CAG AAC GCC AAG GGC CGC TTC CTG AAG ATC GCC GAG
    Asp Val Lys Gln Asn Ala Lys Gly Arg Phe Leu Lys Ile Ala Glu

271 GTG GGC GCG GGC GGC AAC AAG AGC CGC CTT ACT CTC TCC ATG TCA
    Val Gly Ala Gly Gly Asn Lys Ser Arg Leu Thr Leu Ser Met Ser

316 GTG GCC GTG GAG TTC CGC GAC TAC CTG GGC GAC TTC ATC GAG CAC
    Val Ala Val Glu Phe Arg Asp Tyr Leu Gly Asp Phe Ile Glu His

361 TAC GCG CAG CTG GGC CCC AGC CAG CCG CCG GAC CTG GCC CAG GCG
    Tyr Ala Gln Leu Gly Pro Ser Gln Pro Pro Asp Leu Ala Gln Ala

406 CAG GAC GAG CCG CGC CGG GCG CTC AAA AGC GAG TTC CTG GTG CGC
    Gln Asp Glu Pro Arg Arg Ala Leu Lys Ser Glu Phe Leu Val Arg

451 GAG AAC CGC AAG TAC TAC ATG GAT CTC AAG GAG AAC CAG CGC GGC
    Glu Asn Arg Lys Tyr Tyr Met Asp Leu Lys Glu Asn Gln Arg Gly

496 CGC TTC CTG CGC ATC CGC CAG ACG GTC AAC CGG GGG CCT GGC CTG
    Arg Phe Leu Arg Ile Arg Gln Thr Val Asn Arg Gly Pro Gly Leu

541 GGC TCC ACG CAG GGC CAG ACC ATT GCG CTG CCC GCG CAG GGG CTC
    Gly Ser Thr Gln Gly Gln Thr Ile Ala Leu Pro Ala Gln Gly Leu

586 ATC GAG TTC CGT GAC GCT CTG GCC AAG CTC ATC GAC GAC TAC GGA
    Ile Glu Phe Arg Asp Ala Leu Ala Lys Leu Ile Asp Asp Tyr Gly

631 GTG GAG GAG GAG CCG GCC GAG CTG CCC GAG GGC ACC TCC TTG ACT
    Val Glu Glu Glu Pro Ala Glu Leu Pro Glu Gly Thr Ser Leu Thr

676 GTG GAC AAC AAG CGC TTC TTC TTC GAT GTG GGC TCC AAC AAG TAC
    Val Asp Asn Lys Arg Phe Phe Phe Asp Val Gly Ser Asn Lys Tyr

721 GGC GTG TTT ATG CGA GTG AGC GAG GTG AAG CCC ACC TAT CGC AAC
    Gly Val Phe Met Arg Val Ser Glu Val Lys Pro Thr Tyr Arg Asn

766 TCC ATC ACC GTC CGC TAC AAG GTG TGG GCC AAG TTC GGA CAC ACC
    Ser Ile Thr Val Arg Tyr Lys Val Trp Ala Lys Phe Gly His Thr

811 TTC TGC AAG TAC TCG GAG GAG ATG AAG AAG ATT CAA GAG AAG CAG
    Phe Cys Lys Tyr Ser Glu Glu Met Lys Lys Ile Gln Glu Lys Gln

856 AGG GAG AAG CGG GCT GCC TGT GAG CAG CTT CAC CAG CAG CAA CAG
    Arg Glu Lys Arg Ala Ala Cys Glu Gln Leu His Gln Gln Gln Gln

901 CAG CAG CAG GAG GAG ACC GCC GCT GCC ACT CTG CTA CTG CAG GGT
    Gln Gln Gln Glu Glu Thr Ala Ala Ala Thr Leu Leu Leu Gln Gly

946 GAG GAA GAA GGG GAA GAA GAT TGATCAAACAGAATGAAACCCCCACACACAC
    Glu Glu Glu Gly Glu Glu Asp End

998 ACACATGCATACACACACACACAGCCACACACAGAAAATATACTGTAAAGAAAGA

1057 GAGAAAATAAAAAGTTAAAAAGTTAAAAA
```

FIG. 2. Nucleotide sequence of Purα. The nucleotide sequence shown is derived from phage clones λAB6 and λHE1. The amino acid sequence of the open reading frame is indicated beneath the nucleotide sequence. Numbering begins with the first methionine. The line beneath the sequence at base −9 indicates the first base of the λAB6 clone. Singly underlined amino acid residues in boldface type indicate class I repeats. Doubly underlined amino acid residues in boldface type indicate class II repeats.

representative 5′ to the methionine codon at position 1, as determined by computer analysis. Third, the ATG codon is surrounded by sequence features typical of initiation methionines in eukaryotic cells (34). For example, for 85% of bona fide start codons, the nucleotide at position −3 is an A, and in our sequence it is an A. Similarly, the nucleotide at position +4 is usually a G for start codons, and in our case it is a G. Start codons are preceded by two to four C residues from positions −5 to −1, and in our case there are two such C residues. Fourth, RACE reveals that the vast majority of detected 5′ ends of Pur mRNA molecules do not extend significantly beyond the 5′ end shown in Fig. 2. RACE results are described in detail below. The sequence AATAAA, beginning at nucleotide 1062 in Fig. 2, represents a polyadenylation and cleavage signal (see reference 14), consistent with the notion that the 3′ terminus at nucleotide 1085 in Fig. 2 represents the polyadenylation site of the mRNA molecule.

**Modular features of Pur protein sequence.** The deduced amino acid sequence of Purα reveals a modular repeat structure unique among known DNA-binding proteins. There are three repeats of a 23-amino-acid motif (class I repeats) interspersed with two repeats of a 26-amino-acid motif (class II repeats). The class I repeats are shown by single underlining in the sequence of Fig. 2, and the class II repeats are shown by double underlining. While the sequence between these repeats is not conserved, the distance between the class I repeats is highly regular. The repeats themselves are not identical but preserve a number of strictly conserved amino acids at fixed distances along the repeats, indicated by solid boxes in Fig. 3, and a high percentage of conservatively substituted amino acids, indicated by dotted-line boxes in Fig. 3. The aromatic content of these repeats, and the possibility that they are involved in DNA binding, is considered further in the Discussion. The repeat sequences are present in a broad variety of human mRNA molecules, as indicated by hybridization analyses presented below. In addition, at least one other human protein, Purβ, possesses a version of the class I repeat similar, but not identical, to the repeats in Purα, as shown in Fig. 3. A partial sequence of Purβ is presented in a later section. The results indicate a family of protein molecules related by virtue of their possession of a strongly conserved type of repeat module. Modules are organized differently in individual protein molecules.

In addition to repeat modules, Purα contains several structural features of potential importance, denoted in Fig. 4A. Near the amino-terminal end of Purα there is a prominent sequence of 18 glycine residues broken only by a single serine residue. Similar glycine stretches are present in proteins serving a wide variety of functions, including helix-destabilizing proteins (28). Carboxy terminal to all the repeat

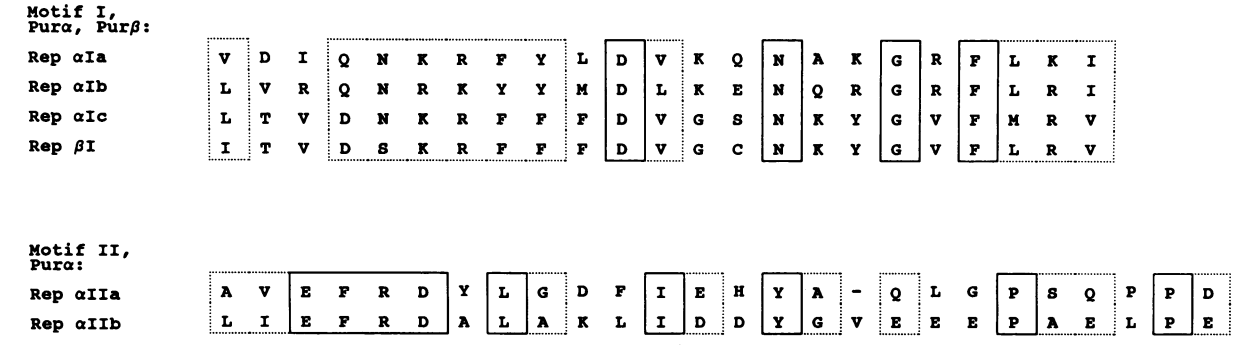


FIG. 3. Two repeat motifs in Purα and Purβ. The class I repeat motifs (three from Purα and one from Purβ) are aligned at the top, and the class II repeat motifs (two from Purα) are aligned at the bottom. Solid boxes indicate identical amino acid residues, and dotted boxes indicate conservative changes.

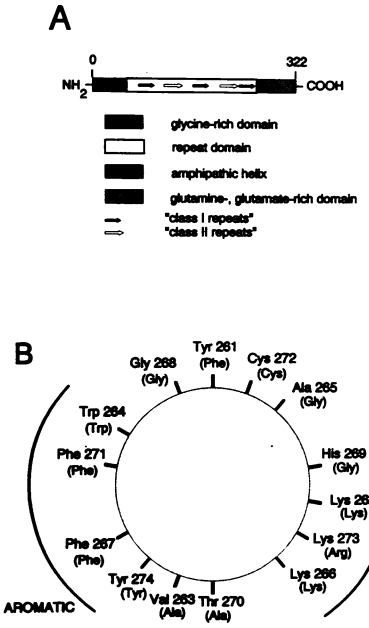


FIG. 4. Arrangement of amino acid sequence domains in Purα. (A) Domain structure of Purα. The glycine-rich and glutamine-glutamate-rich domains both contain 50% or more of those respective amino acids. The class I repeats and the class II repeats are described in Fig. 3. (B) Axial view of the predicted amphipathic helix of Purα. Basic and aromatic faces are indicated. Numbers indicate the position of each amino acid residue in Purα. Bracketed residues indicate the amino acids occupying the equivalent positions in the homologous region of Purβ.

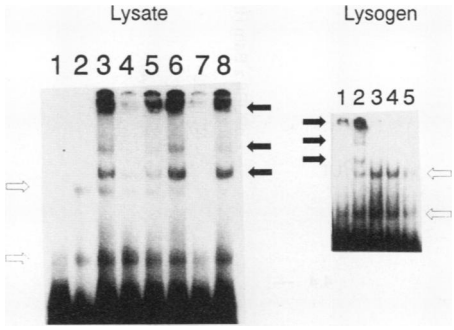


FIG. 5. Gel shift assay of protein extracts from λAB clones. Gel shift assays with labeled MF0677 probe were performed as described in Materials and Methods. Left: Assay of protein extracts prepared from lysates of λAB4. Lane 1 displays probe in the absence of protein. Lane 2 displays probe in the presence of a control protein extract prepared from Y1090 cells infected with λovalb, a chicken ovalbumin gene clone in λgt11. Lane 3 displays probe in the presence of protein extract from Y1090 cells infected with λAB4. Lane 4 displays signal with fivefold dilution of λAB4 extract. Lanes 5 through 8 represent reactions similar to that of lane 3, but with cold competitor added: lane 5, 5-fold excess of MF0677; lane 6, 5-fold excess of poly(A); lane 7, 20-fold excess of MF0677; lane 8, 20-fold excess of poly(A). Solid arrows indicate bands specific to the clone, and open arrows indicate bands present normally in *E. coli*. Right: Assay of lysogens derived from λAB6. Lanes 1 and 2 display signal from IPTG-induced lysogen in either the presence (lane 1) or the absence (lane 2) of a fivefold excess of MF0677. Lane 3 displays signal from uninduced lysogen. Lane 4 displays signal from IPTG-induced Y1090. Lane 5 indicates signal from uninduced Y1090.

modules, there is a region (residues 261 through 274) of potential α-helix (11, 37) upon which the amino acid side chains confer a strongly amphipathic character. The amphipathic helix is ordered with opposing basic and aromatic side chains, as presented in the helical wheel of Fig. 4B. Similar amphipathic helices are present in several DNA-binding proteins thought to play a role in transcriptional activation (43). The carboxy terminus of the Purα molecule consists of a glutamine-glutamate-rich domain. The entire sequence from residue 276 through 321 is 50% glutamine and glutamate residues. There is one sequence of seven consecutive glutamine residues, and near the carboxy terminus there is a sequence of five glutamate residues broken by a single glycine. Glutamine-rich domains have been implicated as transcriptional activation regions in several DNA-binding proteins (12). At the border between the amphipathic helix and the glutamine-glutamate-rich domain, there is the motif Ser-Glu-Glu-Met (residues 275 through 278). The serine in this motif is a potential phosphorylation site for casein kinase II (33), although it is not known whether the motif serves this function in Purα.

**Specificity of single-stranded DNA binding by Purα.** Protein extracts derived from either phage lysates (Fig. 5, left) or lysogens (Fig. 5, right) of λAB5 and λAB6 display the presence of three bands in gel shift assays (solid arrows), using labeled MF0677 as a probe, which are not present in controls. A control lysate was obtained from Y1090 cells infected with λovalb, a chicken ovalbumin clone (Clontech). This lysate showed no *PUR*-binding activity (Fig. 5, left, lane 2). The bands in the experiment with phage lysates (Fig. 5, left) are inhibited by a 20-fold excess of MF0677 (lanes 5 and 7), but not by a 20-fold excess of poly(A) oligonucleotide (lanes 6 and 8), demonstrating specificity in DNA binding by the fusion proteins. In the experiment with lysogen (Fig. 5, right), the clone-specific bands only appear after induction by IPTG (lane 2), indicating that they result from the fusion of the open reading frame indicated for λAB6 in Fig. 2 to the β-galactosidase gene. The band nearest the top of the gel (uppermost arrow in both experiments) is most likely the intact fusion protein since that molecule would be approximately 140 kDa in size and would migrate slowly. The two more rapidly migrating bands generated by each clone (lower solid arrows) are most likely proteolytic products of this larger molecule. There are at least two lower bands in each gel lane that are contributed by *E. coli* (open arrows in Fig. 5), and these proteins are not induced by IPTG (Fig. 5, right, lanes 4 and 5). The *E. coli* proteins also possess some specific affinity for the Pur element.

The specificity of single-stranded DNA binding by Pur was examined in detail by constructing a series of mutated versions of the binding sequence, MF0677, which were used in competition experiments with labeled MF0677 (Table 1). As nucleotides in the single-stranded MF0677 are progressively replaced by thymidine residues, the ability of the altered oligonucleotides to compete is reduced. The oligonucleotide MH0677, which contains GGAGGG, retains the ability to compete. Comparison of this oligonucleotide with ME0677 suggests that the sequence GGAGG is the minimal requirement for binding. GGT repeats near the 5′ end of MF0677 strongly enhance binding. The adenosine residues at the 3′ end of MF0677 are not required for binding, as evidenced by the effective competition with oligonucleotide MC0677, despite the fact that these residues are part of the consensus distribution *PUR* element. These results agree well with the binding site inferred from methylation interfer-

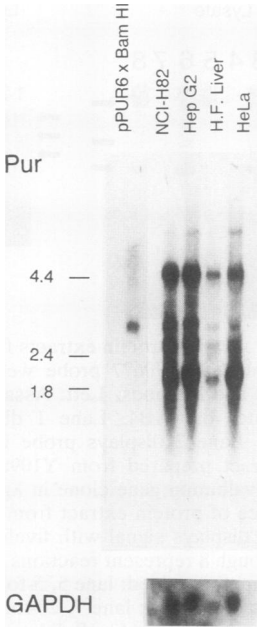


FIG. 6. Blot hybridization of human mRNAs with a *purα* probe. Poly(A)$^+$ RNA prepared from tissue or culture cells was subjected to electrophoresis and blotted as described in Materials and Methods. Membranes were probed with labeled Purα cDNA. Lanes: pPUR6 × BamHI, pPUR6 DNA digested with *Bam*HI; NCI-H82, lung tumor cell line mRNA; HepG2, liver hepatoma mRNA; H.F. liver, human fetal liver mRNA; HeLa, HeLa cell mRNA. The lower box displays the results of hybridization of the same membrane to a glyceraldehyde phosphate dehydrogenase probe as a loading control. Numbers on left show size in kilobases.

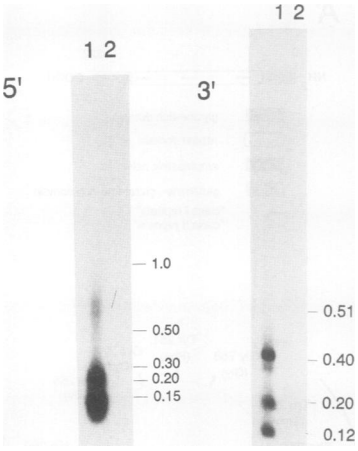


FIG. 7. 5′ and 3′ RACE extension of *purα* cDNA. Products of the RACE reactions outlined in Materials and Methods were subjected to electrophoresis in 1.5% agarose gels containing 40 mM Tris-acetate and 1.0 mM EDTA (pH 8.5), Southern blotted to GeneScreen Plus membranes, and hybridized to a *purα* probe. In each case, numbers on the right indicate molecular size markers in kilobases. Lanes 1 indicate the reaction containing both primers. Lanes 2 represent a control reaction in which one primer was omitted. Primer PDT-01 (TATCTGCAGTTTTTTTTTTTTTTTT) was used to anneal to the poly(A) tail generated for both 5′ and 3′ RACE. Left: 5′ RACE. *pur*-specific primers used were EX-270 (CTCGGCGATCTTCAGGAA), corresponding to nucleotides 270 to 253, for the first amplification reaction, and EX-174 (TTCTAAG CTTCGTCTCGTGCTGCAGCCC), corresponding to nucleotides 174 to 157 plus a *Hin*dIII linker, for the second amplification reaction. For lane 2, primer EX-174 was omitted. Right: 3′ RACE. *pur*-specific primers used were EX-695 (TCTTCGATGTGGGCT CCAAC), corresponding to nucleotides 695 to 714, for the first amplification reaction, and EX-990 (ACACACACACACATGC ATAC), corresponding to nucleotides 990 to 1009, for the second amplification reaction. For lane 2, primer PDT-01 was omitted.

ence studies of Pur activity in HeLa cell nuclear extracts (6). In those experiments, no contacts were seen among the 3′ adenosine residues. In addition, protein contacts were observed with several guanosine residues in MF0677, all of which are also important in the present mutational study. For example, the central G residue of a G triplet makes a prominent contact in methylation interference studies, and it is essential for binding by the cloned and expressed Pur protein (Table 1, compare oligonucleotides MF0677, ME0677, and MG0677). The hamster *dhfr* element (DR3529) possesses the GAGGG motif but does not include the GGT repeats, so it competes somewhat less effectively than does MJ0677 for the human Pur protein. As is true of the HeLa nuclear extract, no binding to Pur is seen with poly(A) or poly(G). These results are strong evidence that the cloned Pur protein possesses the same DNA-binding specificity as the protein identified in HeLa nuclear extracts.

**Multiple human mRNA species are homologous to Purα.** Northern blot analyses were performed on mRNAs isolated from several human cell lines and tissues. A restriction fragment probe of Purα cDNA was chosen to avoid the polyglycine sequence near the amino terminus of the Purα protein. This was done to minimize cross-hybridization since many proteins are known to possess such a glycine sequence. Analyses of mRNAs from human fetal liver tissue, HeLa cells, NCI-H82 lung tumor cells, and HepG2 hepatoma cells result in a similar pattern of multiple transcripts for each tissue or cell type (Fig. 6). The level of each transcript appears to vary little from one cell type to the next. The major transcripts are 2.0 and 5 kb in length, while minor ones occur at 2.7 and 3.2 kb. The overall intensity of bands from the fetal liver sample is lower than that from other cell types, but the intensity of the control GAPDH band for this sample is also lower. This most likely reflects the high level of liver-specific transcripts in the sample. Multiple mRNA species hybridizing to the *purα* probe may be the result of alternate splicing of a single gene or represent transcripts from a number of related genes.

To determine the length of mRNA encoding Purα, we amplified 3′ and 5′ mRNA ends by the PCR RACE techniques described in Materials and Methods. Briefly, these use one PCR primer specific for the Purα sequence and another hybridizing to a homopolymeric tail formed at either the 3′ or 5′ end of a reverse transcript of the mRNA. The results of 3′ RACE are shown in Fig. 7 (right). Three primary bands can be seen at about 420, 200, and 120 nucleotides, indicating that there are three transcripts hybridizing to the Purα primer, that there are three major splicing products of a Purα transcript, or that there are three major 3′-termination sites for a Purα transcript. Further studies of genomic *pur* genes will distinguish among these possibilities. The results of 5′ RACE are shown in Fig. 7 (left). Two primary bands can be seen at 50 and 250 nucleotides, and a diffuse range of much less intense bands can be seen at 500 to 700 nucleotides. The smallest band represents artifactual stopping of either reverse transcriptase or *Taq* polymerase in the G-rich polyglycine sequence near the 5′ end of Purα mRNA.

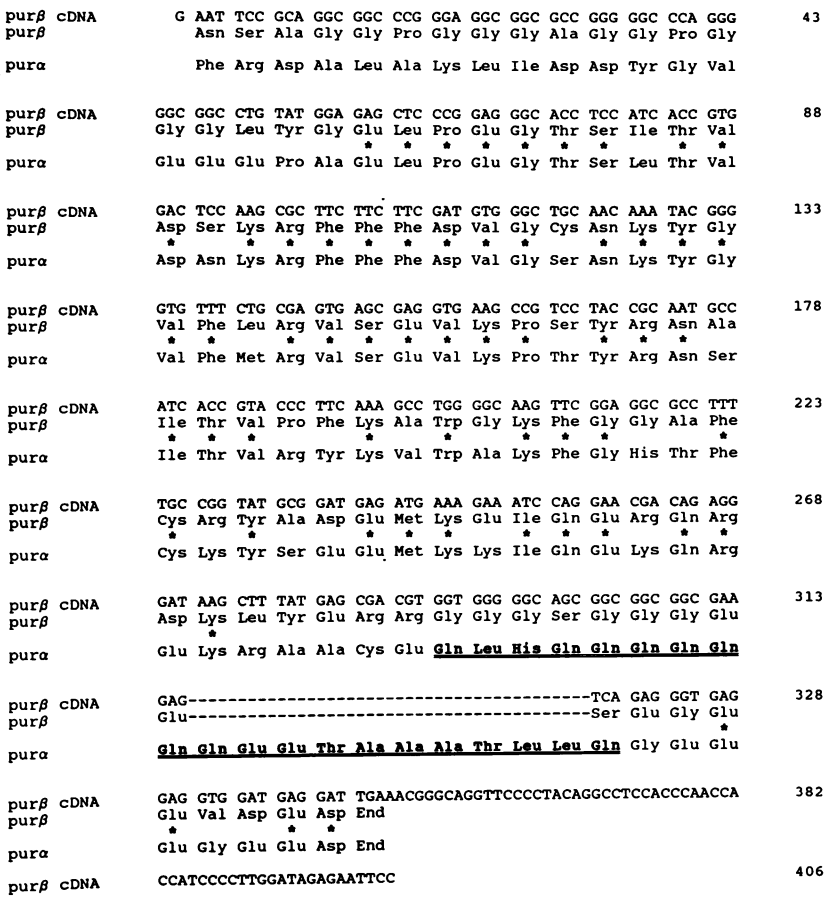


FIG. 8. cDNA sequence encoding the carboxy terminus of the Purβ protein and comparison of the deduced amino acid sequence with the corresponding region of Purα. The DNA sequence of clone λHE4 is presented. Asterisks indicate amino acids homologous between Purβ and Purα. Dotted lines indicate a gap in regions of homology. Double underlining marks the glutamine-rich domain of Purα.

The larger major band corresponds with the 5′ end of the sequence shown in Fig. 2 and could represent the 5′ end of the Purα transcript. The diffuse smear at 500 to 700 nucleotides could conceivably represent longer Purα transcripts. It is more likely, however, based on their low level of production, that they are derived from transcripts related to Purα but with only partial homology to the Purα primer. These 3′ and 5′ RACE results place the size of the Purα mRNA in the range of 1.6 to 2.1 kb. Both the size of the major Purα mRNA and the diversity of homologous species seen are consistent with results obtained by Northern hybridization.

**Screening of a HeLa cell library reveals at least one additional protein with a Pur repeat module.** Probing $2 \times 10^8$ PFU of a HeLa cell λZAP II cDNA library with the $^{32}$P-labeled 777-bp *Pst*I fragment of the Purα cDNA (described in Materials and Methods) yielded several hybridizing clones. One of these is a new clone of Purα, λHE1, which extends the sequence an additional 59 bp at the 5′ end (Fig. 2). Another of these hybridizing clones, λHE4, contains a cDNA with a sequence similar, but not identical, to that of Purα. The protein encoded by this cDNA is designated Purβ. A partial cDNA sequence composing the carboxy terminus of Purβ is presented in Fig. 8. The Purβ protein possesses a copy of the class I repeat module described in Fig. 3 that is followed by an amphipathic helix with considerable sequence similarity to that of Purα (Fig. 3 and 4). However, the 3′ end of Purβ lacks the glutamine-rich domain present at that position in Purα, as shown by the sequence comparison in Fig. 8. There is 80.6% nucleotide homology between the *purα* and *purβ* genes over 216 bp corresponding to the position of maximum protein homology. Over this stretch of 72 codons, there is 75% amino acid homology.

## DISCUSSION

The mechanism by which a protein may achieve sequence specificity in its binding to single-stranded DNA may differ considerably from mechanisms thus far outlined for binding to double-stranded DNA. The interactions of single-stranded-DNA-binding proteins from several bacteriophages, particularly the gene 5 products of bacteriophages Fd, Ike, Pf1, and Pf3 and the gene 32 product of bacteriophage T4, with their substrates have been the subject of considerable study (1, 2, 7, 15, 38, 41, 55). For the Fd and Pf1 gene 5 proteins, chemical modification experiments have established that several tyrosine residues and basic residues are essential for binding to DNA (2, 41). In addition, a phenylalanine residue can become cross-linked to one of the essential tyrosine residues of the Pf1 gene product, suggesting it is also located within the DNA-binding cleft of the protein (41). These studies have led to a hypothesis that these proteins interact with single-stranded DNA by intercalation of the side chains of phenylalanine and tyrosine residues between DNA bases, the aromatic side chains thus becoming involved in base

stacking, while basic residues engage in electrostatic interactions with the phosphates of the backbone. This model may apply to a wide range of single-stranded-DNA-binding proteins. Phenylalanine 469 and lysine 470 of the DBP of adenovirus are known through site-directed mutagenesis studies to be required for single-stranded DNA binding by this protein (40, 45). A number of workers (24, 42, 58) have proposed a motif common to nine prokaryotic and viral single-stranded-DNA-binding proteins which is formed by a series of aromatic and basic residues. It is, therefore, of interest to note that the three class I repeats of Purα, and the copy of the similar module in Purβ, each contain three highly conserved phenylalanine or tyrosine residues and a preponderance of basic residues (Fig. 3). Hydrophilicity profiles of Purα predict that each copy of the class I motif will be on the surface, particularly the first two repeats, raising the possibility that each may be well positioned to form base-stacking interactions with the DNA. There are reports of sequence-specific single-stranded-DNA-binding proteins from mammalian cells that bind sequences rich in cytosine and thymidine residues (17, 20). This raises the possibility that these proteins could cooperate with Purα to stabilize the complementary strand of the DNA and thereby to maintain locally unwound regions of DNA.

The 3′ and 5′ ends of the sequence presented in Fig. 2 agree very well with major bands observed upon 3′ and 5′ RACE amplification of mRNA ends. However, in the absence of an in-frame termination codon in the cDNAs for Purα, the possibility remains that the cloned cDNAs do not contain the entire coding region. In any case, the domain of Purα that is essential for binding single-stranded DNA is clearly contained in the cloned cDNAs. The specificity of the protein expressed from the cloned gene agrees very well with that of the activity observed in the HeLa cell nuclear extract (6).

The predicted sequence of the Purα protein includes several features common to nucleic acid-binding proteins. Substantial glycine-rich regions have been identified in helix-destabilizing proteins and RNA-binding proteins (23, 27, 28, 53). The region from Leu-54 to Leu-75 in Purα, which is moderately helical, contains leucine or isoleucine residues every seventh amino acid residue, raising the possibility of forming a leucine zipper. The sequence from Lys-203 to Lys-229 (KLIDDYGVEEEPAELPEGTSLTVDNK) is typical of PEST (proline-glutamate-serine-threonine) sequences, which are regular features of proteins that are rapidly turned over in the cell (47). It is notable that Purα does not possess a consensus motif (K/R-G-F/Y-A/G-F/Y-V-X-F/Y) commonly found in RNA-binding proteins and thought to be essential for such binding (see reference 23).

Considerable evidence now indicates that DNA replication initiates within discrete zones in mammalian genomes (3, 8, 25, 29, 56, 57). Although size estimates for such initiation zones vary from approximately 0.5 to 30 kb for various loci, the implication is strong that sequences within these zones serve some regulatory function in initiation. There is no direct indication as yet that *cis*-acting sequences act as the binding sites for factors that control initiation, as has been established for several prokaryotic and viral systems. Regions of bent DNA are highly conserved features of replication origins throughout both prokaryotes and eukaryotes, and it has been proposed that they act as binding sites for initiator proteins (16). The binding of Pur to regions of stable bending in two mammalian initiation zones, those in c-*myc* and *dhfr* loci, renders this protein a strong candidate for such a *trans*-acting factor.

Several investigators have recently proposed the involvement of sequence-specific single-stranded-DNA-binding proteins in mammalian recombination (17, 20, 22). The single-stranded-DNA-binding protein RP-A is believed to play a role in simian virus 40 DNA replication and in human DNA recombination (18). Interestingly, it has been recognized that GGN repeats occur commonly in sequences known to function as recombination hot spots, including the breakpoint of oncogene bcl2 translocations (35), the human minisatellite core sequence (32), and the mouse retrotransposon LTR-IS element (17). The D2 element involved in V(D)J recombination in mouse thymocytes contains the motif GGAGGGA, a minimal binding element for Purα (48).

Glutamine-rich regions and amphipathic helices such as those in the carboxy terminus of Purα (Fig. 6) are common features of the transactivation domains of many transcription factors (12, 31, 43). The potential involvement of Purα in replication most certainly does not preclude its involvement in transcription or recombination. In the human c-*myc* locus, the *PUR* element is located in a region previously reported to contain positively acting transcriptional control elements (26). Binding sites for Purα appear within the promoter regions of several mammalian genes (6). A number of factors are involved in both replication and transcription in prokaryotic, lower eukaryotic, and viral systems. For example, the cellular NF-1 protein is involved in transcriptional regulation and also interacts specifically with DNA polymerase during preinitiation of adenovirus DNA replication (9). The yeast ARS-binding factor ABF1 is involved in transcriptional activation, transcriptional silencing, and ARS activation (46). When ABF1 binding sites are replaced by binding sites for other transcription factors, these factors can substitute for ABF1 in ARS activation provided they also possess a transcriptional activation domain (39).

Our mRNA analyses and cDNA cloning results imply that there exists a family of proteins containing Pur class I repeat modules. This type of modular repeat organization is presently unique among known DNA-binding proteins. If the class I modules are involved in DNA binding, as discussed above, then a further implication is that there exists a family of proteins capable of binding similar, but not necessarily identical, repeated DNA elements.

## ACKNOWLEDGMENTS

We thank Edward B. Ziff and D. Stave Kohtz for useful discussions.

This work was supported by the National Institutes of Health (NCI, CA55219).

## REFERENCES

1. **Anderson, R. A., and J. E. Coleman.** 1975. Physicochemical properties of DNA binding proteins: gene 32 protein of T4 and *Escherichia coli* unwinding protein. Biochemistry **14**:5485–5491.
2. **Anderson, R. A., Y. Nakashima, and J. E. Coleman.** 1975. Chemical modifications of functional residues of fd gene 5 DNA-binding protein. Biochemistry **14**:907–917.
3. **Anochkova, B., and J. L. Hamlin.** 1989. Replication in the amplified dihydrofolate reductase domain in CHO cells may initiate at two distinct sites, one of which is a repetitive sequence element. Mol. Cell. Biol. **9**:532–540.
4. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
5. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
6. **Bergemann, A. D., and E. M. Johnson.** 1992. The HeLa Pur

factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Mol. Cell. Biol. **12**:1257–1265.
7. **Brayer, G. D., and A. McPherson.** 1983. Refined structure of the gene 5′ DNA binding protein from bacteriophage fd. J. Mol. Biol. **169**:565–596.
8. **Burhans, W. C., L. T. Vassilev, M. S. Caddle, N. H. Heintz, and M. L. DePamphilis.** 1990. Identification of an origin of bidirectional DNA replication in mammalian chromosomes. Cell **62**: 955–965.
9. **Chen, M., N. Mermod, and M. S. Horwitz.** 1990. Protein-protein interactions between adenovirus DNA polymerase and nuclear factor I mediate formation of the DNA replication preinitiation complex. J. Biol. Chem. **265**:18634–18642.
10. **Chirgwin, J. J., A. E. Przbyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
11. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of protein conformation. Biochemistry **13**:222–245.
12. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59**:827–836.
13. **Dailey, L., M. S. Caddle, N. Heintz, and N. H. Heintz.** 1990. Purification of RIP60 and RIP100, mammalian proteins with origin-specific DNA-binding and ATP-dependent DNA helicase activities. Mol. Cell. Biol. **10**:6225–6235.
14. **Darnell, J. E., Jr.** 1982. Variety in the level of gene control in eukaryotic cells. Nature (London) **297**:365–371.
15. **de Jong, E. A. M., J. P. M. van Duynhoven, B. J. M. Harmsen, R. N. H. Konings, and C. W. Hilbers.** 1989. Two-dimensional $^1$H nuclear magnetic resonance studies of the gene V-encoded single-stranded-DNA binding protein of the filamentous bacteriophage Ike. J. Mol. Biol. **206**:119–132.
16. **Eckdahl, T. T., and J. N. Anderson.** 1990. Conserved DNA structures in origins of replication. Nucleic Acids Res. **18**:1609–1612.
17. **Edelmann, W., B. Kroger, M. Goller, and I. Horak.** 1989. A recombination hotspot in the LTR of a mouse retrotransposon identified in an in vitro system. Cell **57**:937–946.
18. **Erdile, L. F., W. D. Heyer, R. Kolodner, and T. J. Kelly.** 1991. Characterization of a cDNA encoding the 70 kDa single-stranded DNA-binding subunit of human replication protein A and the role of the protein in DNA replication. J. Biol. Chem. **266**:12090–12098.
19. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.
20. **Flavin, M., and F. Strauss.** 1991. Multiple sequence-specific single-strand-binding proteins for the promoter region of rat albumin gene. DNA Cell Biol. **10**:113–118.
21. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNA's from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**:8998–9002.
22. **Gaillard, C., M. Weber, and F. Strauss.** 1988. A sequence-specific single-strand-binding protein for the late-coding strand of the simian virus 40 region. J. Virol. **62**:2380–2385.
23. **Ge, H., P. Zuo, and J. L. Manley.** 1991. Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell **66**:373–382.
24. **Gutierrez, C., G. Martin, J. M. Sogo, and M. Salas.** 1991. Mechanism of stimulation of DNA replication by bacteriophage $\phi$29 single-stranded DNA-protein p5. J. Biol. Chem. **266**:2104–2111.
25. **Handeli, S., A. Klar, M. Meuth, and H. Cedar.** 1989. Mapping replication units in animal cells. Cell **57**:909–920.
26. **Hay, N., J. M. Bishop, and D. Levens.** 1985. Regulatory elements that modulate expression of human *c-myc*. Genes Dev. **1**:659–671.
27. **Haynes, S. R., G. Raychaudhuri, and A. L. Beyer.** 1990. The *Drosophila* Hrb98DE locus encodes four protein isoforms homologous to the A1 protein of mammalian heterogeneous nuclear ribonucleoprotein complexes. Mol. Cell. Biol. **10**:316–323.
28. **Haynes, S. R., M. L. Rebbert, B. A. Mozer, F. Forquignon, and I. B. Dawid.** 1987. Pen repeat sequences are GGN clusters and encode a glycine-rich domain in a Drosophila cDNA homologous to the rat helix destabilizing protein. Proc. Natl. Acad. Sci. USA **84**:1819–1823.
29. **Heintz, N. H., and J. L. Hamlin.** 1982. An amplified chromosomal sequence that includes the gene for dihydrofolate reductase initiates replication within specific fragments. Proc. Natl. Acad. Sci. USA **79**:4083–4087.
30. **Hofmann, J. F.-X., and S. M. Gasser.** 1991. Identification and purification of a protein that binds the yeast ARS consensus sequence. Cell **64**:951–960.
31. **Hope, I. A., and K. Struhl.** 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell **46**:885–894.
32. **Jeffreys, A. J., V. Wilson, and S. L. Thein.** 1985. Individual-specific fingerprints of human DNA. Nature (London) **316**:76–79.
33. **Kenelly, P. J., and E. G. Krebs.** 1991. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. J. Biol. Chem. **266**:15555–15558.
34. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
35. **Krowczynska, A. M., R. A. Rudders, and T. G. Krontiris.** 1990. The human minisatellite consensus at breakpoints of oncogene translocations. Nucleic Acids Res. **18**:1121–1127.
36. **Leffak, M., and C. D. James.** 1989. Opposite replication polarity of the germ line c-*myc* gene in HeLa cells compared with that of two Burkitt lymphoma cell lines. Mol. Cell. Biol. **9**:586–593.
37. **Levin, M., B. Robson, and J. Garnier.** 1986. An algorithm for secondary structure determination in proteins based on sequence similarity. FEBS Lett. **205**:303–308.
38. **Maeda, K., G. Kneale, A. Tsugita, N. Short, R. Perham, D. Hill, and G. Peterson.** 1982. The DNA-binding protein of Pf1 filamentous bacteriophage: amino acid sequence and structure of the gene. EMBO J. **1**:255–261.
39. **Marahrens, Y., and B. Stillman.** 1992. A yeast chromosomal origin of replication defined by multiple functional elements. Science **255**:817–823.
40. **Neale, G. A., and G. R. Kitchingman.** 1989. Biochemical analysis of adenovirus type 5 DNA-binding protein mutants. J. Biol. Chem. **264**:3153–3159.
41. **Plyte, S. E., and G. G. Kneale.** 1991. The role of tyrosine residues in the DNA-binding site of the Pf1 gene 5 protein. Protein Eng. **4**:553–560.
42. **Prasad, B. B. V., and W. Chiu.** 1987. Sequence comparison of single-stranded DNA binding proteins and its structural implications. J. Mol. Biol. **193**:579–584.
43. **Ptashne, M.** 1988. How eukaryotic transcriptional activators work. Nature (London) **335**:683–689.
44. **Quertermous, T.** 1989. Plating libraries and transfer to filter membranes, p. 6.1.1–6.1.4. *In* F. M. Ausubel, M. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 2. Greene Publishing Associates and Wiley Interscience, New York.
45. **Quinn, C. O., and G. R. Kitchingman.** 1986. Functional analysis of the adenovirus type 5 DNA-binding protein: site-directed mutants which are defective for adeno-associated virus helper activity. J. Virol. **60**:653–661.
46. **Rhode, P. R., S. Elsasser, and J. L. Campbell.** 1992. Role of multifunctional autonomously replicating sequence binding factor I in the initiation of DNA replication and transcriptional control in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12**:1064–1077.
47. **Rogers, S., R. Wells, and M. Rechsteiner.** 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science **234**:364–368.
48. **Roth, D. B., P. B. Nakajima, J. P. Menetski, M. J. Bosma, and M. Gellert.** 1992. V(D)J recombination in mouse thymocytes: double-strand breaks near T cell receptor $\delta$ rearrangement

signals. Cell **69**:41–53.
49. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, vol. 1, 2nd ed., p. 7.43–7.45. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
50. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
51. **Singh, H.** 1989. Detection, purification, and characterization of cDNA clones encoding DNA binding proteins, p. 12.7.1–12.7.10. *In* F. M. Ausubel, M. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 2. Greene Publishing Associates and Wiley Interscience, New York.
52. **Singh, H., J. H. LeBowitz, A. S. Baldwin, Jr., and P. A. Sharp.** 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52**:415–423.
53. **Steinert, P. M., J. W. Mack, B. P. Korge, S. Q. Gan, S. R. Haynes, and S. C. Stevens.** 1991. Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricins and single-stranded RNA binding proteins. Int. J. Biol. Macromol. **13**:130–139.
54. **Strauss, W. M.** 1989. Hybridization with radioactive probes, p. 6.3.1–6.3.6. *In* F. M. Ausubel, M. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 2. Greene Publishing Associates and Wiley Interscience, New York.
55. **van Duynhoven, J. P. M., P. J. M. Folkers, A. P. M. Stassen, B. J. M. Harmsen, R. N. H. Konings, and C. W. Hilbers.** 1990. Structure of the DNA binding wing of the gene V-encoded single-stranded DNA binding protein of the filamentous bacteriophage M13. FEBS Lett. **261**:1–4.
56. **Vassilev, L., and E. M. Johnson.** 1990. An initiation zone of chromosomal DNA replication located upstream of the c-*myc* gene in proliferating HeLa cells. Mol. Cell. Biol. **10**:4899–4904.
57. **Vaughn, J. P., P. A. Dijkwel, and J. L. Hamlin.** 1990. Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain. Cell **61**:1075–1087.
58. **Wang, Y., and J. D. Hall.** 1990. Characterization of a major DNA-binding domain in the herpes simplex virus type 1 DNA-binding protein (ICP8). J. Virol. **64**:2082–2089.
59. **Wobbe, C. R., L. Weissbach, J. R. Borowiec, F. B. Dean, Y. Murakami, P. Bullock, and J. Hurwitz.** 1987. Replication of simian virus 40 origin-containing DNA in vitro with purified proteins. Proc. Natl. Acad. Sci. USA **84**:1834–1838.