Order Number 9510689

Structure and chromosomal localization of the gene encoding Pur-alpha, a sequence-specific single-stranded DNA-binding protein in human and mouse

Ma, Zhi-Wei, Ph.D.

City University of New York, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

STRUCTURE AND CHROMOSOMAL LOCALIZATION OF THE GENE ENCODING PUR-ALPHA, A SEQUENCE-SPECIFIC SINGLE-STRANDED DNA-BINDING PROTEIN IN HUMAN AND MOUSE

b y

ZHI-WEI MA

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1994

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/23/94
Date

Dr. Edward M. Johnson
Chairman of Examining Committee

9/26/94
Date

Dr. Terry A. Krulwich
Executive Officer

Dr. Nathaniel Heintz*
Dr. D. Stave Kohtz
Dr. Natalie W. Tulchin
Dr. Lu-Hai Wang

Supervisory Committee

* The Rockefeller University

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Structure and chromosomal localization of the gene encoding Pur-alpha, a sequence-specific single-stranded DNA-binding protein in human and mouse

by

Zhi-Wei Ma

Advisor: Professor Edward M. Johnson

Pur-alpha is a sequence-specific single-stranded DNA-binding protein with affinity for an element present in several eukaryotic origins of DNA replication and gene regulatory regions. I have cloned and sequenced the cDNAs for human and mouse Pur-alpha (hPur-alpha and mPur-alpha, respectively). There is an extraordinary degree of conservation between hPur-alpha (322 aa) and mPur-alpha (321 aa), and there are only two amino acid (aa) differences between them. One Pur-alpha region of 22 aa, termed the "psycho" motif, possesses significant homology to counterparts in the SV40 large T-antigen, several other transforming proteins of DNA tumor

viruses and certain cellular proteins in yeast and human cells that may also be involved in the initiation of DNA replication.

Multiple mRNA species are homologous to Pur-alpha cDNA with similar pattern in several human cell lines and tissues. I have attempted to distinguish between alternate processing of the same pre-mRNA from a single gene, the presence of a related genes, or both. I have isolated human genomic clones for Pur-alpha and another family member, Pur-beta, which contains cDNA sequence similar to that of Pur-alpha. Based on Southern analysis of human genomic DNA and genomic clones, both human *pur*-alpha and *pur*-beta are demonstrated to be single copy genes. Pur-alpha, expressed as a GST-fusion protein, specifically binds to the hypophosphorylated form of the retinoblastoma protein, Rb, with an affinity at least as high as that of SV40 large T-antigen. Binding of Pur-alpha to Rb alters the conformation of the Pur-alpha complex with its DNA recognition element as seen under standard gel shift conditions.

The *pur*-alpha gene is localized to chromosome 5q31 by fluoresence *in situ* hybridization (FISH) of genomic DNA probes on human metaphase chromosome spreads and by Southern analysis of human/hamster somatic hybrid cells containing single human chromosome. The site 5q31 is frequently altered in nonlymphocytic leukemia and in other cancers. Results of the experiments suggest that Pur-alpha may be involved in the cell cycle control of DNA replication.

DEDICATION

This dissertation is dedicated to my wonderful wife, Hai-Wen, and my parents.

ACKNOWLEDGEMENTS

I am greatly indebted to Dr. Edward M. Johnson for his critical guidance, valuable advice, and constant encouragement throughout my graduate study.

I would like to thank Dr. Ronald E. Gordon and Dr. Saul Puszkin for serving on my advisory committee, and for their helpful suggestions on my study.

I would like to thank Sharon M. Barr, Andrew D. Bergemann, Yan Dai, Li Gan, Chavdar P. Krachmarov, Mechael Kanovsky, and Ernest Savransky for their help in many ways.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Overview of DNA replication

DNA replication is the process by which a cell duplicates its genome. The start sites of replication are called origins of replication. The DNA regions replicated from each single origin are termed replicons. Prokaryotic genomes and eukaryotic mitochondrial DNA genomes have only one replication origin. Eukaryotic genomic DNA initiates replication at mutiple locations in each chromosome and replicates most of chromosomal DNA from origins irregularly spaced at intervals of 50-300 kilobases (Huberman & Riggs, 1968; Edenberg & Huberman, 1975). By far the preponderance of such origins initiate bidirectional DNA synthesis, although unidirectional DNA synthesis has been reported in yeast (Linskens & Huberman, 1988).

Origins of DNA replication

Replication origins in prokaryotes, eukaryotic mitochondria, animal DNA tumor viruses, and lower eukaryotes contain specific sequence elements of functional importance (reviewed by Campbell, 1986; DePamphilis, 1988; Stillman, 1989; Umek et al., 1989). They usually contain two primary components: a core consensus sequence as a recognition element for a replication initiator protein, and a flanking sequence as a readily unwound element, commonly rich in AT content (Campbell, 1986; Umek & Kowalski, 1988; Kowalski &

Eddy, 1989; Umek et al., 1989). Replication origins in eukaryotes may have additional characteristics: they may have other auxillary sequence components involved in regulation of replication initiation efficiency, timing and cell type specificity (DePamphilis, 1988; Walker et al., 1990; Ferguson & Fangman, 1992; Dobbs et al., 1994). Replication initiation is specified by the interaction between these *cis*-acting sequences and *trans*-acting factors, including replication initiator proteins as well as accessory proteins.

Initiation of DNA replication

Initiation of DNA replication at both *E. coli* replication origin *oriC* and simian virus 40 (SV40) replication origin has been characterized in great detail. These origins are frequently used as models to illustrate the structural components and mechanism of DNA replication initiation. The core element at *oriC* is a 9 base pair repetitive sequence, while the core element at the SV40 origin is a GAGGC repeat. Replication initiator proteins DnaA, for *oriC* of *E. coli*, and large T antigen, for the SV40 origin, bind specifically to the core elements in their origins and form multimeric complexes (Bramhill & Korberg, 1988a; Mastrangelo et al., 1989; Borowiec et al., 1990). These protein-DNA interactions either promote or enhance DNA bending (DNA bending will be discussed later in this introduction), leading to unwinding of a flanking sequence near the binding sites. These DNA unwinding elements, termed *DUE*,

are apparently common to all origins, both prokaryotic and eukaryotic. The initiator proteins thus facilitate the entering of accessory proteins e.g., DnaB and DnaC in *E. coli oriC* (Bramhill & Kornberg, 1988a; Borowiec et al., 1990). Following the entry of DnaC-DnaB (DnaB has helicase activity) into the opening at *oriC*, gyrase (topoisomerase II) and single strand DNA binding protein (SSB) further unwind and stabalize the template and expose it to the DNA replication machinery (Bramhill & Kornberg, 1988a; Bramhill & Kornberg, 1988b). The nascent DNA strand synthesized continuously is called the leading strand, while the other nascent DNA strand synthesized discontinuously at each replication fork is called the lagging strand. RNA primers are required for DNA polymerase III at the start of DNA synthesis of the nascent strands. Primase, a specific RNA-synthesizing enzyme, makes short RNA primers followed by DNA polymerase III synthesis of DNA semi-discontinuously in a 5' to 3' direction as bidirectionally growing forks (Bird et al., 1972; Prescott et al., 1972; Kornberg, 1988). In the lagging strand, the RNA primers are repeatedly required to generate the short nascent strands, which are termed Okazaki fragments; the RNA primers are degraded and replaced by DNA polymerase I action with the assistance of RNaseH. The gaps between any two newly-synthesized DNA fragments are linked together by ligase (Kornberg, 1988). The rate of initiation of *oriC* linearly depends on DnaA concentration (Herrick & Bensimon, 1991).

Initiation of DNA replication in yeast

Yeast is frequently employed as a model for the studies of DNA replication in eukaryotes, partly because its chromosomal organization and replication are typical of higher eukaryotes, but its genome is much less complex (Campbell, 1986). Certain genomic DNA fragments from yeast (*S. cerevisiae*)have been characterized as autonomously replicating sequences or *ARS* elements because these sequence elements allow plasmids containing them to replicate autonomously in yeast cells (Stinchcomb et al., 1979). Two-dimensional gel mapping methods have confirmed that *in vivo* replication initiates at *ARS* elements in both yeast 2 μm plasmids and cellular chromosomes (Huberman et al., 1987; Brewer & Fangman, 1987; Linskens & Huberman, 1988; Linskens & Huberman, 1990) There are 400 *ARS*s per cell, but only a fraction of *ARS* elements appear to be used as chromosomal replication origins *in vivo* (Campbell, 1986; Brewer & Fangman, 1988; Huberman et al., 1988; Dubey et al., 1991; Fangman & Brewer 1991). Each *ARS* element contains an 11-bp core consensus sequence (Broach et al., 1982). Any single point mutation in this sequence eliminates *ARS* function (Houten & Newlon, 1990). However, this sequence itself is not sufficient for autonomous replication of plasmids (Kearsey , 1984; celniker et al., 1984; Maine et al., 1984). The region flanking the consensus sequence is also required for replication initiation. The lengths and the sequences in the flanking region

vary with different *ARS*s (Broach et al., 1982; Ceniker et al., 1984; Kearsy, 1984; Brand et al., 1987). The flanking sequence of an *ARS* element contains several important subset elements, each of which may have distinct function (Celniker et al., 1984; Marahrens & Stillman, 1992). When these subset elements (referred to as B1, B2 and B3) are mutated individually, a slight reduction of *ARS1* activity is observed. When two or three of the B elements are simultaneously mutated, *ARS1* function is severely compromised (Marahrens & Stillman, 1992). Since some transcription factors bind to these subset elements, they may contribute to the functional links between replication and transcription.

Initiation of DNA replication in higher eukaryotes

Accumulated data support the notion that unique sequences as origins of replication are also present in higher eukaryotes. The mechanism of replication initiation at a given origin is highly conserved between prokaryotes and higher eukaryotes (for review see Taylor, 1984; DePamphilis, 1988; Laskey et al., 1989; Umek et al., 1989; Diffley & Stillman, 1990; Virshup, 1990; Marraccino et al., 1990; DePamphilis, 1993). However, studies of replication initiation in higher eukaryotic cells have been hampered mainly by the complexity of the genome. For instance, the genome of a human cell is over 200 times as large as that of a yeast cell. Furthermore, studies in higher eukaryotes are complicated by the mutiple coordinated

controls over replication initiation in chromosomal DNA. These controls, which operate at different levels, are as follows: (1) Regulation of initiation may be a primary level of replication control which has evolved from prokaryotes (Diffley & Stillman, 1990; Marraccino et al., 1990). (2) Transcription of an adjacent gene may regulate replication initiation in terms of timing, efficiency and cell type specificity (Goldman et al., 1984; Taylor, 1984; Snyder et al., 1988; DePamphilis, 1988; Heintz, 1992; Heintz et al., 1992). A chromosomal origin of DNA replication is co-localized with the transcriptional enhancer in the human *c-myc* gene zone (Iguchi-Ariga et al., 1988). A common element for both initiation of replication and repression of transcription has been reported recently in yeast (Rivier and Rine, 1992; Foss et al., 1993; Bell et al., 1993). (3) Many replication origins are clustered together and activated synchronously, and different clusters are activated at different time intervals during the synthetic phase (Hand, 1978; Nakamura et al., 1984; Nakamura et al., 1986; Bravo et al., 1987; Dubey & Raman, 1987; Mills et al., 1989; Hutchison & Kill,1989; Nakayasu & Berezney, 1989; Diffley & Stillman, 1990; Cox & Laskey, 1991). This temporal regulation of activation of DNA replication origins can be affected by chromosomal location of the origins. The *cis*-acting element determining origin function can be separated from the *cis*-acting element controling temporal activation of replication origin (Ferguson and Fangman, 1992), although this has only been observed in yeast. (4) Origin usage in embryonic cells and other somatic

cells may be controlled differently (Blumenthal et al., 1973; Callan, 1973; Kriegstein & Hogness, 1974; DePamphilis, 1993). (5) Chromatin structure and nucleoskeleton may play important roles in the organization and function of replication origins. For example, they can determine the topology of the replication origins and origin accessibility to *trans*-acting factors (Simpson, 1990; Jackson, 1990; Alberts, 1990; Hurwitz et al., 1990; Brown et al., 1991; Cook, 1991; DePamphilis, 1993).

Initiation zones for DNA replication in higher eukaryotes

Much effort has been expended in order to identify origins of replication in mammalian chromosomal DNA. Some investigators have sought to functionally identify mammalian DNA fragments that are capable of supporting autonomous episomal replication when the DNA fragments are cloned into plasmids or viral constructs with deficient replication origins and re-introduced into mammalian cells. Others have sought to physically locate and characterize authentic initiation zones in chromosomes. One of the most thoroughly characterized origins is in the dihydrofolate reductase (*dhfr*) gene region in Chinese hamster ovary (CHO) cells. There are three major advantages that have made this localization possible: (1) a CHO-derived cell line (CHOC 400) has been developed in which each cell contains about 1,000 copies of homogeneous *dhfr* domain (135 kb) in its genome (Milbrandt et al., 1981); (2) the *dhfr* domain in CHOC

400 has been shown to be replicated at the very begining of the S phase (Milbrandt et al., 1981); (3) the same replication pattern for the amplified units of the *dhfr* domain suggests that replication of each repeated *dhfr* domain is initiated from the same origin(s) (Heintz & Hamlin, 1982).

A 28 kb initiation zone of DNA replication is first identified approximately 17 kb 3' to the *dhfr* gene by *in vivo* labeling and molecular cloning (Heintz & Hamlin, 1982; Heintz et al., 1983), although other researcheres assert that the initiation zone is located at a broader region by two-dimensional gel methods (Vaughn et al., 1990; Dijkwel et al., 1991). This potential initiation zone is further narrowed down to a 4.3 kb fragment (referred to here as *oridhfr-1*) by *in vivo* labeling followed by restriction digestions (Burhans et al., 1986a; Burhans et al., 1986b). Two distinct origins separated by 22 kb within the previously mapped 28 kb locus are suggested by strand extrusion method and in-gel renaturation analysis. One of them is in agreement with the *oridhfr-1* (Anachkova & Hamlin, 1989; Leu & Hamlin, 1989). The presence of two distinct origins in the *dhfr* region is further supported by a nucleosome segregation pattern mapping method (Handeli et al., 1989). The zone of *oridhfr-1* can serve as a functional replication origin for an origin deficient viral construct in CHO cells (Handeli et al., 1989). A 2.5 kb initiation zone of *oridhfr*-1 is defined (Vassilev et al., 1990) in non-amplified Chinese hamster cells by a polymerase-chain-reaction (PCR) based mapping method developed in our

laboratory (Vassilev & Johnson, 1989). A 450 bp region within this 2.5 kb zone has been identified by an Okazaki fragment polarity mapping method as a center of origin of bidirectional replication (OBR), (Burhans et al., 1990). Analysis in the *oridhfr*-1 reveals a repetitive element, a region of stably bent DNA, two Alu repeats, several elements homology to *ARS* consensus sequence, several repeats of purine-rich elements, and several DNA unwinding elements (Caddle et al., 1990; Leu et al., 1990).

Another well-characterized origin is located upstream of the *c-myc* gene in HeLa cells (a human cervical adenocarcinoma cell line). Fragments from 5' flanking DNA of *c-myc* has been reported to support autonomous replication of plasmids in HL-60 cells (Iguchi-Ariga et al., 1988) and HeLa cells (McWhinney and Leffak, 1990), and to be the chromosomal replication origin based on results of an *in vitro* run-off replication assay (McWhinney and Leffak, 1990). Vassilev and Johnson (1990) were the first to map the *c-myc* initiation zone in chromosomes *in vivo*. They reported an initiation zone of bidirectional replication, which is 2 kb in length, centered approximately 1.5 kb upstream of the P1 promoter of the human *c-myc* gene by the PCR based mapping method. Since the sequence surrounding the initiation zone near the *c-myc* gene is known (Battey et al., 1983), detailed sequence comparison between this origin zone, the *oriDHFR-1* region and other newly-mapped zones are now possible.

In addition to the hamster *dhfr* and the human *c-myc* zones described above, there are three newly-mapped initiation zones of DNA replication in mammalian cells. They are the human β-globin (Kitsberg et al., 1993), mouse adenosine deaminase (Carroll et al., 1993; Virta-Pearlman et al., 1993) and hamster rhodopsin (Gale et al., 1992) gene initiation zones. All five of these initiation zones possess regions of potential, or demonstrated, DNA bending. They also possess prominent A-T rich, potential DNA unwinding elements, *DUE*. In each case, near these potential *DUE*, there is another highly purine-rich element known as *PUR* element (Bergemann and Johnson, 1992; for review see Dobbs et al., 1994), which will be discussed below. No functional elements of a replication origin have been characterized in higher eukaryotes.

Proteins involved in initiation of DNA replication in eukaryotes

The search for replication initiator proteins and regulatory proteins has been carried out mainly by DNase 1 footprinting assays and gel retardation assays with DNA fragments from identified replication origins. A yeast *ARS* consensus sequence binding protein, ACBP, has been identified and purified recently (Kuno et al., 1990; Schmidt et al., 1991; Hofmann & Gasser, 1991). This protein binds specifically to the T-rich single-stranded DNA of the yeast *ARS* consensus sequence. It also binds the double-stranded version and a

number of 9/11 matches to the *ARS* consensus with lower affinity (Kuno et al., 1990; Schmidt et al., 1991; Hofmann & Gasser, 1991). Three additional proteins have been identified to bind to a subset of *ARS* elements. (Eisenberg et al., 1988; Diffley & Stillman, 1988; Francesconi & Eisenberg, 1989; Walker et al., 1990; Buchman & Kornberg,1990). The exact functions of any of these proteins in DNA replication are not clear yet. The yeast *ARS*-binding factor I (ABF1) is involved in transcriptional activation, transcriptional silencing, and *ARS* activation (Rhode et al., 1992). When ABF1 binding sites are replaced by binding sites for other transcription factors, these factors can substitute for ABF1 in *ARS* activation provided that they also possess a transcriptional activation domain (Marahrens and Stillman, 1992). Recently, a multiprotein complex that specifically recognizes the *ARS* consensus sequence (ACS) and one subset element (B1) has been identified and purified from *S. cerevisiae* (Bell and Stillman, 1992). This multiprotein complex is refered to as the origin recognition complex (ORC) and contains eight polypeptides which could not be separated by chromatographic steps. There is a strong correlation between origin function and origin binding by the ORC. It is likely that ORC is an initiator protein complex at yeast replication origins (Bell and Stillman, 1992).

MCM (MiniChromosome Maintenance) are a group of related yeast proteins, such as MCM2 and MCM3, which are involved in the initiation of chromosomal DNA replication in an origin-dependent manner in yeast (for review see Koonin,

1993). BM28, a human MCM2 homolog, is reportedly required for entry of cells into S-phase (Todorov et al., 1994). These cellular proteins possess homologies to SV40 large T-antigen (Ma et al., 1994a), a well-characterized replication initiation protein and transcription factor.

RIP60 is a 60-kDa HeLa cell nuclear protein. It recognizes a specific DNA sequence which is located immediately adjacent to the 3' end of the bent DNA in the hamster *oridhfr-1* region (Dailey et al., 1990). The recognized sequence is primarily comprised of ATT repeats. This protein can enhance DNA bending in the *oridhfr-1* region (Caddle et al., 1990). Studies by electron microscopy and scanning transmission electron microscopy reveals that RIP60 binds ATT repeats at three or more sites in the hamster *oridhfr-1* region. RIP60 binds as dimers or multiples of dimers and forms link structures with the DNA. These link structures render an intervening DNA sequence of 700 bp into supercoiled loop DNA (Mastrangelo et al., 1993).

Replication protein A (RPA) is a single-stranded DNA-binding protein. It is required for replication initiation at the simian virus 40 origin *in vitro* (Wobbe, et al., 1987; Wold and Kelly, 1988; Erdile, et al., 1991). RPA is shown to have preference for the pyrimidine-rich strand of the inverted repeat (IR) within the SV40 replication origin (Kim et al., 1992). No sequence specificity has been reported for DNA binding by RPA. IR factor B (IRF-B), a human nuclear protein of 34 KDa, is another single-stranded DNA-binding protein. It

recognizes the purine-rich strand of the inverted repeat (Carmichael et al., 1993). RPA and IRF-B bind opposite strands of the IR domain within the SV40 replication origin and together they may involve in the regulation of origin activation (Carmichael et al., 1993).

DNA bending and its relevance to DNA replication

Long chains of DNA are quite flexible, while short pieces of DNA are relatively stiff. The shorter the DNA chains are, the more energy is needed to bend them. However, certain DNA sequences are more susceptible to bending than others. Bent DNA fragments are usually detected by their slower mobility in polyacrylamide gel electrophoresis in cold temperature than predicted from their actual size (Wu and Crothers, 1984). They also can be predicted by computer analysis (Eckdahl and Anderson, 1987; Eckdahl and Anderson, 1990). Bent DNA fragments frequently contain blocks of A_{2-6} in a 10-11 base periodicity (Anderson, 1986). This spacing causes successive deflections in the same direction, resulting in macroscopic bending at this site (Eckdahl and Anderson, 1990). However, other sequences, particularly purine-rich sequences, can also contribute to DNA bending (Bergemann and Johnson, 1992).

Sequences of DNA bending have been found to be conserved in origins of DNA replication in plasmids, bacteriophages, bacteria, yeast and mammals (Stenzel et al., 1987; Zahn and Blattner, 1987; Eckdahl and Anderson, 1990;

Caddle et al., 1990; Begermann and Johnson, 1992). It is believed that DNA bending, in general, may facilitate DNA-protein interactions. Many proteins, including histones, bend DNA at the sites they bind. Some proteins, such as SV40 large T-antigen and RIP60, enhance DNA bending (Caddle et al., 1990). These observations support the notion that the structures of DNA bending play an importnant role in DNA replication.

The PUR element was originally identified at a site of prominent DNA bending upstream of human *c-myc* gene

In our laboratory, a 467-bp *Sau*3A DNA fragment at 1.65 kb upstream of the *c-myc* P1 promoter was identified as an intrinsic DNA bending site (Bergemann and Johnson, 1992) by restriction cleavage cyclic permutation assay. This DNA bending site is located within the previously mapped initiation zone of DNA replication (Vassilev and Johnson, 1990). A protein, designated Purα, from HeLa cell nuclear extracts has been identified. It recognizes a specific sequence element (designated the Purα element or *PUR* element) near the center of the bent fragment. The Purα protein preferentially binds single-stranded rather than double-stranded *PUR* element (Bergemann and Johnson, 1992). The preferred single-stranded *PUR* element is a purine-rich 24-mer nucleotides, GGAGGTGGTGGAGGGAGAGAAAAG (MFO677). This sequence is

conserved in the initiation zone of hamster *oridhfr-1* (Caddle et al., 1990), the initiation zone of mouse adenosine deaminase (Virta-Pearlman et al., 1993), the 5' flanking sequence of the *N-myc* gene in human (Stanton et al., 1986) and mouse (DePinho et al., 1986), and several origins of replication in lower eukaryotes (for references see Bergemann & Johnson, 1992). There are two transcription regulatory sites near the *c-myc PUR* element. A *CTF/NF-1* consensus binding element is present 200 to 300 bp downstream of the *PUR* element (Siebenlist et al., 1984). Another element, termed the far upstream element (*FUSE*), is located between the *CTF/NF*-1 and *PUR* elements of the *c-myc* gene zone in human chromosomes. A 70-kDa protein factor, termed FBP, is reported to be a sequence-specific, single-stranded DNA-binding protein. It stimulates c-*myc* gene transcription through binding to the *FUSE* (Duncan et al., 1994).

The human Pur protein

The human Purα protein is a sequence-specific single-stranded DNA-binding protein with affinity for an element present in several eukaryotic origins of DNA replication and gene regulatory regions (Bergemann et al., 1992). It was identified as a HeLa cell nuclear protein, with a molecular weight of 29 kDa (Bergemann and Johnson, 1992) as estimated by UV cross-linking technique. It was first cloned by screening expression libraries for proteins with affinity for the labeled,

single-stranded *PUR* element (Bergemann et al., 1992). One of the positive clones, λAB6, was subcloned and sequenced. The obtained 1.08 kb cDNA sequence contains an open reading frame, designated Purα, which has 322 amino acids in length. 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 cell extracts. The cDNA sequence was used as the probe for subsequent library screening and nucleic acid hybridization. One of the positive clone, obtained by DNA-DNA hybridization, contains a cDNA with a sequence similar, but not identical, to that of Purα. The protein encoded by this cDNA is designated Purβ. There is 80.6% nucleotide homology between the purα and purβ genes over 216 bp corresponding to the position of maximum protein homology (Bergemann et al., 1992).

Available evidence, as following, suggests that the Purα protein is involved in initiation of DNA replication. (1). The Purα-binding element, *PUR*, is found near *DUE* in all five mapped initiation zones of DNA replication in mammalian chromosomes. (2). The Purα protein contains homologous domain to SV40 large T-antigen, including the Rb-binding domain (Ma et al., 1994a). (3). The Purα protein contains a domain homologous to a group of DNA-dependent ATPases, such as DnaA of *E. coli* (Koonin, 1993; Ma et al., 1994a). Proteins sharing this homology include yeast MCM2, MCM3 (Yang et al., 1991), human nuclear protein BM28 (Todorov et al., 1994; Ma et al., 1994a) and FBP (Duncan et al., 1994). Nevertheless, at

this time we do not know the exact function of the Purα protein.

CHAPTER 2

SPECIFIC AIMS

The overall goal of my project is to provide the structural groundwork for identifying the functions of Purα. I have intended to clone and sequence Purα genes and to identify genomic clones for these genes in human and mouse. I sought to help elucidate potential cell-cycle regulatory aspects of Purα function through its interaction with the retinoblastoma protein, Rb. Finally, I sought to ascertain the chromosomal location of the Purα gene in human cells. Specifically, I have:

1. cloned the cDNA of Purα, a 29 kDa replication origin binding protein in human cells;
2. determined the size of the complete Purα cDNA and its expression in different tissues using Northern blot and RACE procedures;
3. documented the protein interaction between Purα and retinoblastoma protein, Rb;
4. cloned and sequenced the mouse Purα cDNA and documented the extraordinary degree of conservation between the human and the mouse Purα;
5. isolated the genomic clones of human *purα* and *purβ* genes;
6. determined the chromosomal localization of the human *purα* gene.

CHAPTER 3

MATERIALS AND METHODS

1. Stock solutions

NZYM

1000 ml of NZYM contains 5g of NaCl, 5g of yeast extract, 10g of Casamino acids, 2g of $MgSO_4 \cdot 7H_2O$, 3 ml of 1N NaOH.

DNA prehybridization solution

300 ml of the solution contains 30 ml of 0.5 M TES, pH7.0, 45 ml of 20xSSC (1xSSC is 150 mM NaCl, 15 mM Sodium Citrate, pH 7.0), 30 ml of 10% SDS, 60 ml 50% dextran sulfate, 7.5 ml of NaH_2PO_4, 7.5 ml of Na_2HPO_4, 120 ml of H_2O, and 100 μg/ml sonicated salmon sperm DNA.

RNA prehybridization solution

300 ml of the solution contains 30 ml of 0.5 M TES, pH7.0, 30 ml of 20 x SSC, 30 ml of 10% SDS, 60 ml 50% dextran sulfate, 7.5 ml of NaH_2PO_4, 7.5 ml of Na_2HPO_4, 135 ml of H_2O, 100 μg/ml sonicated salmon sperm DNA and 100 μg/ml yeast RNA.

Library screening solution

500 ml of the solution contains 50 ml of 50 x Denhardt's solution [1% Ficoll (Type 400, Pharmacia), 1%

polyvinylpyrrolidone, 1% bovine serum albumin (Fraction V, Sigma)], 150 ml of 20 x SSC, 0.1% SDS, 1 ml of 0.5 M EDTA, and 100 μg/ml sonicated salmon sperm DNA.

2. Oligonucleotides used

PDT-01:	TATCTGCAGTTTTTTTTTTTTTTTT	*Pst*I
EX-270:	CTCGGCGATCTTCAGGAA	
EX-174:	TTCTAAGCTTCGTCTCGTGCTGCAGCCC	*Hind*III
EX-695:	TCTTCGATGTGGGCTCCAAC	
EX-990:	ACACACACACACATGCATAC	
MFO677:	GGAGGTGGTGGAGGGAGAGAAAAG	
NOTR:	CATCTCCTTCGAGTACTTGCAGAA	
NOTF:	GAACCAGCGCGGCCGCTTCCTGCG	

3. Cell lines and cell culture

The cell culture media, fetal bovine serum, and antibiotics used in cell culture were purchsed from GIBCO. All media were supplemented with 10% fetal bovine serum, 100 units/ml penicillin G, and 100 μg/ml streptomycin.

Chinese hamster ovary (CHO) cells were grown in monolayer flasks. All other cell lines used were grown in suspension in one litre Corning spinner flasks. Hela cells were grown in Dulbeco's modified eagle medium, $HepG_2$ cells were grown in Joklik modified eagle medium, and human small cell lung carcinoma cells (NCI-H82) were grown in RPMI 1640 medium. CHO cells were grown in Dulbeco's modified eagle

meduim, supplemented with 100 μM hypoxanthine, 16 μM thymidine, and 0.3 mM proline (Sigma).

Cells were counted with a hemocytometer. Cells were collected by centrifugation at 1,000 x g for 5 minutes and washed with 1 x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4).

4. Preparation of competent *E. coli* cells

One hundred ml of Luria-Bertani (LB) broth was inoculated with 1 ml of overnight culture of JM 83 cells. After shaking at 37°C for three hours, the cells were centrifuged at 4,000 rpm in a Sorval GS-3 rotor for 10 minutes. The cell pellet was suspended in 10 ml of 0.1 M $CaCl_2$ (ice cold, sterilized by passing through a 0.22 μm filter) and incubated on ice for 20 minutes. The cells were centrifuged at 4,000 rpm in a SS-35 rotor for 5 minutes. The cell pellet was resuspended in 4 ml of ice cold 0.1 M $CaCl_2$. Two ml of 50% glycerol was added and mixed with the cell suspension. The cells were dispensed into 200 ml aliquotes on dry ice/ethanol, and kept at -70°C.

5. Transformation of competent *E. coli* cells

The competent E. coli cells (JM83 or XL-1) were thawed on ice. One to ten ng of circular pBluescript II KS (-) plasmid DNA construct was gently mixed with 100 μl of the competent cells on ice. The cells were incubated on ice for 30 minutes and

heat-shocked at 42°C for 90 seconds. After incubation on ice for additional 2 minutes, the cells were transfered to 500 µl of prewarmed LB medium, and incubated at 37°C for 1 hour with 185 rpm shaking. One hundred µl of the cell suspension was spread on a LB agar plate (90-mm petri dish) containing 50 µg/ml ampicillin. The plate was prewarmed and spread with 50 µl of 2% 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal), and 5 µl of 100 mM isopropylthio-β-D-galactoside (IPTG) for blue/white colony selection. The plate was then incubated in 37°C overnight.

6. RNA preparation

Human fetal liver tissue blocks from several 19 week abortuses were kindly supplied by Drs. Sunkara Rao, Renata Dische and Stave Kohtz (Mount Sinai School of Medicine, New York). The human liver tissue blocks were frozen in liquid nitrogen and ground in a Biopulverizer homogenizer. The poly(A)$^+$ RNA from both harvested cells and tissue blocks was extracted and purified using RNA extraction and mRNA purification kits (Pharmacia) according to the manufacturer's instructions. RNA samples were passaged twice through oligo dT-cellulose columns to ensure greater than 90% of the eluted RNA was poly(A)$^+$ RNA. The RNA concentration of each sample was determined through measurement of its optical absorbance at 260 nm.

7. Preparation of genomic DNA

Fifty to one hundred million cultured cells (Hela cells and CHO cells) were lysed in 10 mM ethylenediaminetetraacetate (EDTA), 50 mM Tris, pH 8.0, and 0.5% sodium dodecyl sulfate (SDS). Pancreatic ribonuclease was added to a final concentration of 20 μg/ml. Samples were incubated in 37°C for 2 hours. Then proteinase K was added to a final concentration of 0.1 mg/ml. The samples were incubated in 50°C for 5-12 hours. DNA was extracted 3 times with equal volume of phenol-chloroform-isoamyl alcohol (25:24:1), once with chloroform, and once with ether. After addition of 0.1 volume of 3.0 M sodium acetate, pH 5.2, 3 volume of undiluted ethanol (-20°C), high molecular weight genomic DNA was spooled out with glass rods, gently rinsed with 70% ethanol, and dissolved in TE buffer (1 mM EDTA, 10 mM Tris, pH 8.0). DNA concentration was determined by spectrophotometry.

Human placenta genomic DNA was either a gift from Dr. Tom M. Fasy (Mount Sinai School of Medicine, New York) or purchased from Sigma.

Genomic DNA of Human/Chinese hamster somatic hybrid cells containing only human chromosome 5 (repository#: NA10114) or only human chromosome 6 (repository#: NA11580) were purchased from NIGMS Human Genetic Mutant Cell Repository.

8. Preparation of phage DNA

One hundred ml of NZYM (5 gm of NaCl, 5 gm of yeast extract, 10 gm of casamino acids, 2 gm of $MgSO_4.7H_2O$, and 3 ml of 1N of NaOH for 1,000ml preparation) was inoculated with 1 ml of fresh overnight culture of host bacterial cells (Y1090 for EMBL3, C600-Hfi for λgt10). After 2.5 hours incubation at 37°C with shaking (the OD_{600} for the bacterial suspension usually reaches 0.5), $2x10^9$ phage particles (PFU) of interest were added. The incubation was continued for additional 6 to 8 hours until the cells were completed lysed. Two ml of chloroform was added to the flask, with constant shaking for 15 minutes at 37°C.

The phage lysate was centrifuged at 10,500 rpm in a Beckman JA-14 rotor for 10 minutes to get rid of the cell debris. DNase I and RNase were added to the supernatant to a final concentration of 1 μg/ml. After incubation at 37°C for one hour, the sample was cooled on ice, and added 50 ml of 30% polyethylene glycol (PEG 8000) in 3 M NaCl, and kept at 4 °C overnight.

The phage particles were precipitated by centrifugation at 12,000 rpm for 30 minutes. The phage pellet was suspended in 5 ml SM buffer (0.1 M NaCl, 10 mM $MgSO_4$), and extracted with chloroform twice to get rid of the trace PEG. The phage suspension was then adjusted to 0.5% SDS and 20 mM EDTA. Proteinase K was added to the sample to a final concentration of 100 μg/ml. The sample was incubated at 50°C for 2 hours,

cooled on ice, and extracted with phenol-chloroform twice and chloroform once. After addition of 0.1 volume of 3.0 M sodium acetate, pH 5.2, and 3 volume of undiluted ethanol (-20^{0}C), the phage DNA was spooled out with glass rods, gently rinsed with 70% ethanol, and dissolved in 0.5 ml of TE buffer, pH 8.0.

9. Preparation of plasmid DNA

The boiling method (Sambrook, et al., 1989b;) was used for mini-preparation of plasmid DNA.

Midi- or maxi-preparation of plasmid DNA was carried out by using QIAGEN silicagel column according to the manufacturer's instructions, except that RNase A was added after potassium acetate precipitation.

10. Preparation of restriction fragments

Restriction fragments in 10 to 20 µl were separated on agarose gels with TAE (4 mM Tris-acetate, pH 8.5, 1 mM EDTA) running buffer. At the end of the run, the gels were stained in 0.5 µg/ml ethidium bromide for 30 minutes and photograghed. DNA fragments were purified from the agarose gels by using either of the following methods.

A. Purification using GeneClean II kit

DNA fragments of interest were cut from the gels. The gel slices were transferred to individual microcentrifuge tubes and weighed. Three volumes (according to the weigh of the gel slice) of 5 M NaI stock solution was added, and the tubes were incubated at 50°C for 5 minutes with intermittent swirling until the gel slices were completely dissolved. Each tube was added 5 µl of the Glassmilk (Bio 101) and put on ice for 5 minutes with intermittent shanking. DNA samples were pelleted with the Glassmilk by centrifugation at full speed for 10 seconds. The samples were washed three times with 700 µl of ice cold New Wash (Bio 101) by spinning at full speed for 10 seconds each time, and then resuspended in 5 µl of TE buffer or distilled water. After incubated at 50°C for 3-5 minutes, the Glassmilk was pelleted by centrifugation at full speed for 10 minutes, and the supernatant containing the DNA sample was transferred to another tube for subsequent use.

B. Purification using NA45 DEAE membrane

After electrophoretic separation in agarose gels, strips of NA45 membrane (Schleicher & Schuell) were placed in incisions just ahead of or behind the DNA bands of interest. Electrophoresis were continued until DNA bands were completely transfered to the membranes, as judged by ethidium bromide fluorescence using long wave UV. Each strip

of membrane ahead of DNA bands was freed of residual agarose by rinsing in 1 x TAE, and then submerged in 200 µl of DEAE elution buffer (1.0 M NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) in a microcentrifuge tube. The tube was incubated at 55-68°C for 15-30 minutes, and then the buffer was collected. After adding 500 µl of undiluted ethanol, the purified DNA was precipitated, washed with 70% ethanol, and redissolved in H_2O for subsequent use.

11. Ligation of restriction fragments

Five µg of plasmid vector DNA (pBluescript II phagemaid) was digested with restriction enzyme(s) in 100 µl reaction volume. After the digestion was complete, 2 µl of alkaline phosphatase (BioLabs, 10 units/µl) was added and the reaction was allowed to continue for additional one hour at 37°C. The plasmid vector DNA was purified from agarose gel as described in preparation of restriction fragments. Equal mole of the vector DNA (usually 100 ng) and the restriction fragment to be cloned were dissolved in 8 µl of H_2O. One µl of 10 x Ligase buffer and 40 units of DNA ligase (BioLabs) were added. The reaction was performed at 14°C overnight.

12. Labeling of DNA probes

A. Random primer labeling

The labeling reactions were performed using the Random Priming System I (BioLabs) according to the manufacturer's instructions. Forty ng of DNA (restriction fragment purified from agarose gel, as described in preparation of restriction fragments) was dissolved in 34 µl of distilled H_2O, and denatured in boiling water for 5 minutes. The centrifuge tube was quickly placed on ice, and 5 µl 10 x labeling buffer (BioLabs, random hexadeoxyribonucleotides included), 5 µl dNTP mixture (1.7 µl of dATP, dTTP, and dGTP, each in a concentration of 0.5 mM), 5 µl [α-^{32}P] dCTP(3,000Ci/mmol, 50 mCi), and 1 µl DNA polymerase I-Klenow Fragment (5 units) were added. The tube was placed at 37°C for 3 hours. Then the reaction mix was loaded onto a Sephadex G-50 Quick Spin Column (Phamacia), centrifuged at 2,000 rpm for 2 minutes, and collected in a screw-capped microcentrifuge tube. The DNA probe was denatured by heating in a boiling water bath for 5 minutes before it was used directly for hybrydization.

B. 5' End labeling of oligonucleotides

Six hundred ng of oligonucleotide, or 1 µg of DNA/RNA size markers in 1 µl was mixed with 1 µl of 10 x kinase buffer, 1 µl of T4 polynucleotide kinase (BioLabs), and 7 µl of [γ-^{32}P]ATP. The reaction was performed at 37°C for 2 hours. The reaction mix was diluted to 50 µl with H_2O, loaded onto a Bio-Spin®6 Chromatography Column (BIO-RAD, Cat.# 732-6002) in a collection tube, and centrifuged at 1,100 x g for 5 minutes.

The purified oligonucleotide solution was diluted to 1,000 µl with H_2O for subsequent use in gel shift assays. The purified DNA/RNA size markers were used for Southern/Northern analysis.

C. Biotin labeling of DNA probes

The labeling reaction was performed by using the Non-Isotopic Probe Labeling Kit (Oncor®, Cat.# S4089-KIT). The manufacturer's instruction manual was followed. Two µg of LGT1 or LGT6 DNA dissolved in 70 µl of H_2O was mixed with 20 µl of nucleotide mixture (with biotin-labeled nucleotides) and 10 µl of nick translation enzyme mixture on ice. The reaction was performed at 15°C for 75 minutes, and then 10 µl of 0.5 M EDTA was added to stop the reaction. The probes were subsequently used for chromosome *in situ* hybridization.

13. Northern analysis

Ten µg of poly(A)$^+$ RNA was used for each sample. RNA bands were separated in 1.2% agarose gel containing 2.2% formaldehyde (Sambrook et al., 1989a) at 3 volts/cm for 3.5 hours. At the end of the run, the lanes containing the RNA size markers were cut from the gels and stained in 1.5 µg/ml ethidium bromide for 30 minutes, destained in distilled water for 2 hours or overnight, and photograghed with alignment of UV transmissible rulers. The gels were rinsed with distilled

water and submerged in distilled water at 4°C overnight. The gels were submerged in 10 x SSC (1 x SSC is 150 mM NaCl, 15 mM sodium citrate, pH 7.0) for 45 minutes before the RNA was transfered to GeneScreen Plus membranes over a sponge for 8 hours with 10 x SSC. Membranes were exposed to UV light for 2 minutes and then baked at 80°C in a vacuum oven for 2 hours. Membranes were prehybridized for 3 hours at 70.5°C in RNA prehybridization solution (2 x SSC , 50 mM TES, pH 7.0, 25 mM Na_2HPO_4, 25 mM NaH_2PO_4, 1% SDS, 10% dextran sulphate, 100 μg/ml sonicated salmon sperm DNA and 100 μg/ml yeast RNA). Hybridization to the probe was performed in the same solution at 70.5°C for 17 hours. The 777-bp *Pst*1 fragment from the λAB6 (see Fig. 1) was used as the probe, which was purified from 1% agarose gels by NA45 membrane and labeled using the random priming kit (BioLabs). After hybridization, the membranes were washed twice for 5 minutes each in 2 x SSC and 1% SDS at 70.5°C and once for 30 minutes in 0.1 x SSC and 1% SDS at 70.5°C. Membranes were then rinsed in 2 x SSC and autoradiographed on Kodak XAR5 films.

14. Genomic Southern analysis

Ten μg of genomic DNA [from HeLa cells (human), C2 myoblast cells (mouse), and human/Chinese hamster somatic hybrid cells] in 150 μl of reaction volume was digested 5 hours with each of the following restriction enzymes: *EcoR*I, *Hind*III, *Pst*I, *Sac*I, and *BamH*I. DNA from each reaction was purified by

extraction with phenol-chloroform twice, and precipitated with ethanol. The DNA digests were separated on 0.8% agarose gels. The gels were stained in 0.5 μg/ml ethedium bromide for 30 minutes and photograghed with alignment of UV transmissible rulers. The gels were treated in 0.2 N HCl for 10 minutes, in 0.5 N NaOH and 1.5 M NaCl for 45 minutes, and then in 0.5 M Tris, pH 7.5 and 1.5 M NaCl for 40 minutes. DNA was transferred to GeneScreen Plus membranes (Du Pont) with 10 x SSC. The membranes were exposed to UV light for 2 minutes and prehybridized in the DNA prehybridization solution (3 x SSC, 50 mM TES, pH 7.0, 25 mM Na_2HPO_4, 25 mM NaH_2PO_4, 1% SDS, 10% dextran sulfate, and 100 μg/ml of sonicated sermon sperm DNA) at 65°C for 2 hr. The *Pst*I fragment from the Pur cDNA AB6 was radioactively labeled as the probe by using the Random Priming kit (BioLabs). Hybridization was performed in the same solution at 65°C for 17 hr. The membrane was then washed twice in 2 x SSC and 1% SDS at 65°C each for 5 minutes, twice in 0.1 x SSC and 1% SDS at 65°C each for 30 minutes. After being rinsed in 0.1 x SSC, the membrane was autoradiographed on Kodak XAR5 film.

15. Rapid amplification of cDNA ends (RACE)

HepG2 cell poly(A)$^+$ RNA used here was prepared as decribed in RNA preparation above. For details of the primers used here, see Table 1 and Figure 1.

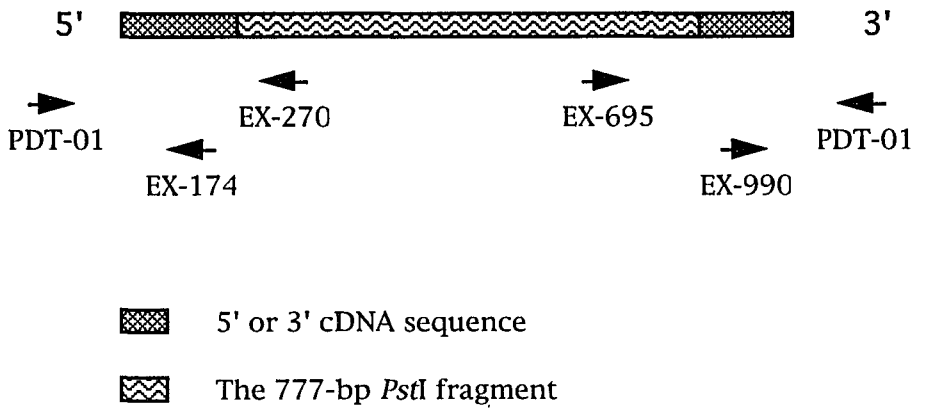

Fig. 1. Schematic diagram of human Purα cDNA and application of RACE techniques to Purα. The horizonal straight bar refers to the the human Purα cDNA sequence from clones PZT6 (see the sequence analysis of human Purα cDNA in the Results). The human Purα cDNA sequence contains the 777-base pair *Pst*I fragment used as the probe in Southern and Northern hybridization and library screening. Arrows indicate the orientation and the nucleotide positions of the beneath oligonucleotide primers. The primers were designated according to the Purα cDNA sequence, except that PDT-01 referred to the primer $d(T)_{17}$ with a *Pst*I recognition sequence at its 5' end. Primers with leftward arrows or rightward arrows refer to their sequences complementary to the plus strand or minus strand of the Purα cDNA sequence, respectively. The primer sequences are presented on page 23. Primers were combined in PCRs as described in Materials and Methods.

1) Synthesis of the first strand cDNA for 3' RACE of Purα

One μg of HepG2 cell poly(A)$^+$ RNA 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 20 μmoles of primer PDT-01 and 200 units of reverse transcriptase from the SuperScript Preamplification System (BRL) according to the manufacturer's instructions, except that 20 units of RNasin (Promega) was used in the reaction.

2) The first amplification of 3' cDNA end of Purα

Polymerase chain reaction (PCR) buffer was supplied by Promega, and 0.2 mM of each dNTP was added to 1 x PCR buffer (this 1 x PCR buffer was used in most of PCRs in this thesis and each PCR was performed in 100 μl of reaction volume with 2.5 units of Taq polymerase, except indicated otherwise). PCR was carried out with 1 μl of the cDNA pool above and two primers (100 μM of both PDT-01 and EX-695). The PCR profile was follows: for 25 cycles, denaturation at 94°C for 1 minute, annealing at 54°C for 2 minutes, and extension at 72°C for 5 minutes.

3) The second amplification of 3' cDNA end of Purα

PCR was performed with 1 μl of the first amplification pool and two primers (100 μM of both PDT-1 and EX-990), and was carried out for 20 cycles with the same profile as described above.

4) Synthesis of the first strand cDNA for 5' RACE of Purα

One μg of HepG2 poly(A)$^+$ RNA was reverse transcribed as described above, except for the substitution of 20 μmoles of primer EX-270 for PDT-01. The cDNA was purified from excess primer EX-270 by passing through a Centricon 100 microconcentrator (Amicon).

5) 3' tailing of the first strand cDNA for 5' RACE of Purα

Poly(A) tail was added to the first strand cDNA synthesized above at the 3' end by terminal deoxynucleotidyl transferase (TdT; BRL) in the presence of 100 μM of dATP at 37°C for 30 minutes. The cDNA was heated at 70°C for 2 minutes, precipitated with ethanol, and resuspended in 20 μl of TE buffer, pH 7.5.

6) The first amplification of 5' cDNA end of Purα

The cDNA pool of Pur above was amplified with primers EX-270 and PDT-01 using 2.5 units of Pyrococcus DNA polymerase (Pfu; Stratagene) in 100 μl. The PCR profile was following: denaturation at 97.5°C for 1 minute, annealing at 54 C for 2 minutes, extension at 74°C for 3 minutes. After 15 cycles, additional 2.5 units of Pfu was added and the reaction was allowed for additional 15 cycles.

7) The second amplification of 5' cDNA end of Purα

One μl of the PCR products above was used for the second amplification at the same conditions as described above, except that primer EX-270 was substituted by internal primer EX-174 and the total of 20 reaction cycles was continuous without addition of extra Pfu.

8) Southern analysis of the RACE products of Purα

Ten μl from each of the RACE products above were separated on 1.5% agarose gel, and transfered to GeneScreen Plus membrane. Hybridization and washing conditions were the

same as described in Southern analysis. The probe used was the whole insert (1,080 bp) of the λAB6, purified from 1% agarose gel using GeneClean II kit (Bio 101), and random-primer labeled. Membranes were autoradiographed on Kodak XAR5 films.

16. Screening a human cDNA library

A human testis cDNA library (Clontech, prepared in vector λgt10) was screened for clones of Pur. This library was plated on four 23 x 23 cm plates with *Escherichia coli* Y1090 at a density of 3.75 x 10^5 PFU per plate. These plates were incubated at 37°C for 8 hours and then transferred to 4°C overnight.

These plates were then overlaid with nitrocellulose filters, one minutes for the first replica, three minutes for the second replica. After being lifted from the plates, the filters were immerged in 0.5 N NaOH and 1.5 M NaCl once for 5 minutes; 0.5 M Tris-HCl, pH 7.5 and 1.5 M NaCl twice for 5 minutes; then rinsed in 2 x SSC for 5 minutes. The filters were dried, and baked at 80°C in a vacuum oven for 1.5 hours.

The filters were prehybridized at 65°C for 2 hours in library screening solution [0.1% Ficoll (Type 400, Pharmacia), 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin (Fraction V, Sigma), 6 x SSC, 0.1% SDS, 1 mM EDTA, and 100 μg/ml sonicated salmon sperm DNA]. Hybridization to the probe was performed in the same solution at 65°C for 19 hours. The

1095-bp *EcoR*1 insert from the λAB6 (see Fig. 1) was used as the probe, which was purified from 1% agarose gels by NA45 paper (Schleicher & Schuell) and labeled using the random priming kit (BioLabs). After hybridization, filters were washed twice for 5 min each in 2 x SSC and 0.1% SDS at 65°C and once for 30 min at the same conditions. Filters were then rinsed in 2 x SSC and autoradiographed on Kodak XAR5 films.

17. Screening a mouse cDNA library

A mouse fetal heart cDNA library (cloned in vecter λZAP II) was screened for mouse Pur clones. The 777-bp *Pst*I fragment from human Purα cDNA was used as the probe, as previously described in screening libraries by DNA-DNA hybridization (Bergemann, et al., 1992).

18. Screening a human genomic DNA library

A human placenta genomic library was screened for genomic clones of Pur. This library is made by cloning *Sau3A*I partial digests at the *BamH*I sites in EMBL3 vectors which can be excised from the recombinants at externally flanked *Sal*I sites (Clontech). This library was plated on five 23 x 23 cm plates with *Escherichia coli* Y1090 at a density of 3.2 x 10^5 PFU per plate. Plates were incubated at 37°C overnight and then transferred to 4 °C.

The subsequent treatments were performed at exactly the same condition as described in screening human cDNA library, but with the 777-bp *Pst* I fragment from human Purα cDNA as the probe.

19. DNA sequencing

Inserts in λgt10, λZAP II and EMBL3 were cloned or subcloned into pBluescript phagemaid vector.

Five μg of the plasmid DNA in 80 μl H_2O was mixed with 20 μl of 1 M NaOH + 1 mM EDTA, and incubated at 37°C for 30 minutes. The mixture was neutralized by adding 10 μl of 3 M sodium acetate (pH 5.2), and DNA was precipitated with 300 μl of ethanol (-20°C, over 30 minutes).

Sequencing of double-stranded recombinant plasmid DNA was performed by using SEQUENASE Version 2.0 and TAQuence Version 2.0 (United States Biochemical) with the dideoxy termination technique (Sanger et al., 1977). The manufacturer's protocols for sequencing reaction were strictly followed, except that 1 μl of [^{35}S]dATP was used in each set of reactions. Sequencing of double-stranded PCR products was performed by using sequencing grade *Taq* DNA polymerase (*fmol* DNA Sequencing System, Promega). The reaction mixture was separated in 6% polyacrylamide/urea sequencing gels [containing 7.6 M urea and TBE (90 mM Tris-borate, pH 8.3, 2 mM EDTA)] at 90 watts. Sequence analysis was conducted by using the IBI Pustell program.

20. Chromosome *in situ* hybridization

Preparation of metaphase-chromosome spreads from a healthy individual, chromosome *in situ* hybridization, and digital microscopic image analysis were performed as previously described (Najfeld et al., 1992).

21. Gel shift assays

Gel shift assays were performed as described peviously (Ausubel et al., 1989). Ten µl of binding reaction mixture contained 0.15 ng of MFO677 oligonucleotide probe (described in 5' end labeling of oligonucleotide), 0-100 ng of purified Pur-GST fusion protein, 0-50 ng of purified retinoblastoma protein (R* $P56^{Rb}$), 10% glycerol, 50 mM KCl, 1 mM MgCl2, 1mM Ethyleneglycol-bis-(b-aminoethyl ether) N,N,N'N'-tetraacetic acid (EGTA, Sigma), 25 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES, pH 7.5), 1 mM Dithiothreitol (DTT), 200 µg/ml bovine serum albumin, and 0.5 µg poly(dI-dC). Binding was carried out at 30°C for 15 minutes. Binding reaction mixture was immediately loaded on 6% polyacrylamide gels that had been pre-run for 2 hours. Electrophoresis was conducted in TBE buffer at 150 volts for 3.5 hours at 4°C. Gels were dried onto Schleicher & Schuell GB002 paper and autoradiographed on Kodak XAR5 film.

CHAPTER 4

RESULTS

1. Characterization of human Purα cDNA

A. Isolation of a Purα clone from a λgt10 library by DNA-DNA hybridization and sequence analysis of Purα cDNA

The human Purα protein was first identified in HeLa cell nuclear extracts as a sequence-specific single-stranded DNA-binding protein (Bergemann et al., 1992). It was first cloned by screening expression libraries for proteins with affinity for the labeled, single-stranded *PUR* element (Bergemann et al., 1992). One of the positive clones, λAB6, was subcloned into pUC19 vector (this construct was referred to as pPUR6) and sequenced. The obtained cDNA sequence was 1.08 kb. The protein encoded by λAB6 possesses single-stranded DNA-binding capacity with a specificity similar to that of the Purα protein observed in HeLa cell nuclear extracts. It was indicated from Northern blot analysis of several human cell lines and tissues that the transcripts hybridized to the Purα cDNA probe was larger than 1.08 kb (detailed Northern blot analysis will be described later in this chapter). Subsequent human cDNA library screening was performed using the known Purα cDNA as probe in order to obtain overlapping cDNA sequences.

Screening of 1.5 x 10^6 plaques from a human testis cDNA library in λgt10 with labeled 1.08 kb cDNA insert from λAB6, as described in Materials and Methods, yielded one positive plaque (designated LZT6) after three rounds of consecutive

screening. The cDNA insert in LZT6 was subcloned into pBluescript phagemaid vector (this construct was referred to as PZT6) and sequenced. The nucleotide sequence of the *pur*α gene insert in PZT6 is presented in Figure 2.

```
-78                 AAA GCA GCG GCG GCT GAG GCG ACT GAG GCG GCG

-45  GGC GGA GCG GCA GGC GGC GGC GGC GCG GCA GCG GAG CGC AGC ATC

  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 CCC TAC AAG GTG TGG GCC AAG TTC GGA CAC ACC
     Ser Ile Thr Val Pro 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 TGA TCA AAC AGA ATG AAA CCC CCA
     Glu Glu Glu Gly Glu Glu Asp End

991  CAC ACA CAC ACA TGC ATA CAC ACA CAC ACA CAG CCA CAC ACA CAG

1036 AAA ATA TAC TGT AAA GAA AGA GAG AAA TAA AAA GTT AAA AAG TTA

1081 AAA AAA AAA AAA
```

Figure 2. Nucleotide sequence and deduced amino acid sequence of Purα from clone PZT6. The nucleotide sequence shown is derived from clone PZT6. The amino acid sequence of the open reading frame is indicated beneath the nucleotide sequence. Numbering begins with the first methionine. Two *Pst*I sites within the nucleotide sequence are underlined.

The obtained 1,170-bp cDNA sequence contains a open reading frame of 322 amino acids in length, 5' untranslated sequence of 78-bp, and 3' untranslated sequence of 126 bp. The 3' untranslated sequence contains a polyadenylation signal. The 5' untranslated sequence does not contain an in-frame stop codon at this point, as it may happen naturally. Two *Pst*I sites

within the nucleotide sequence are specifically underlined because the 777-bp *Pst*I fragment (also see Fig. 1) is frequently used as the Purα cDNA probe in the subsequent experiments.

B. Features of Purα protein sequence

```
M A D R D S G S E Q G G A A L G S G G S L G H P G  25
S G S G S G G G G G G G G G G S G G G G G G A  50
P G G L Q H E T Q E L A S K R V D I Q N K R P Y L  75
D V L Q N A L G R F L K I A E V G A G G N K S R L  100
T L S M S V A V E F R D Y L G D F I E H Y A Q L G  125
P S Q P P D L A Q A Q D E P R R A L K S E F L V R  150
E N R K Y Y M D L K E N Q R G R F L R I R Q T V N  175
R G P G L G S T Q G Q T I A L P A Q G L I E F R D  200
A L A K L I D D Y G V E E E P A E L P E G T S L T  225
V D N K R F F F D V G S N K Y G V F M R V S E V K  250
P T Y R N S I T V P Y K V W A K F G H T F C K Y S  275
E E M K K I Q E K Q R E K R A A C E Q L H Q Q Q Q  300
Q Q Q E E T A A A T L L L Q G E E E G E E D  322
```

Fig. 3. The deduced amino acid sequence of human Purα and its features. The glycine-rich sequence is marked by a single thin underline. The glutamine-glutamate-rich sequence is marked by a dashed underline. The amphipathic helix region is marked by a single thick underline. Numbering begins with the first methionine.

The deduced amino acid sequence of Purα contains several structural features of potential importance. Near the amino-terminal end of Purα there is a prominent sequence of 18 glycine residues (residues 31 through 49, marked by the single thin underline in Fig. 3) broken only by a single serine residue. Similar glycine stretches are present in proteins serving a wide variety of functions, including helix-destabilizing proteins (Haynes et al., 1987). The carboxyl terminus of the Purα molecule consists of a glutamine-glutamate-rich domain (marked by the dashed underline in Fig. 3). The entire sequence from residue 276 through 321 is 50% glutamine and glutamate residues. Glutamine-rich domains have been implicated as transcription activation domains in several DNA-binding proteins (Courey et al., 1989). Amino terminal right to the glutamine-glutamate-rich domain, there is a region (residues 261 through 274, marked by the single thick underline in Fig. 3) of potential α-helix upon which the amino acid side chains confer a strongly amphipathic character. The amphipathic helix is ordered with opposing basic and aromatic side chains. Similar amphipathic helices are present in several DNA-binding proteins thought to play a role in transcription activation (Ptashne, 1988). 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 (Kenelly and Krebs, 1991), although it is not known whether the motif serves this function in Purα.

C. Multiple human mRNA species are homologous to Purα

In order to determine the size of the complete Purα cDNA and to reveal the tissue distribution of its expression, Northern blot analyses were performed on poly(A)$^+$ RNA isolated from several human cell lines and tissues. A *Pst*I restriction fragment of human Purα cDNA, as indicated in Fig. 2, was chosen as the probe to avoid the highly GC-rich sequence (83% GC content) near the 5' terminus of the cDNA (Fig. 1 and Fig. 2). Analysis of mRNA from HeLa cells, NCI-H82 lung tumor cells, HepG2 hepatocarcinoma cells, and human fetal liver tissue revealed multiple transcripts with similar pattern for each cell line and tissue studied (Fig. 4). The level of each transcript appears to vary little from one cell type to the other. The major transcripts are 2.0 and 5 kb in length, while minor ones occur at 2.7 and 3.2 kb. Multiple transcripts hybridized to the Purα cDNA probe may be the result of alternate processing of the same pre-mRNA from a single gene, or represent transcripts from a family of related genes, or both. This question will be addressed later in this chapter.

Membranes were hybridized to mouse 28S RNA cDNA (a gift from D. Stave Kohtz, Mount Sinai School of Medicine, New York) probe at the same conditions to conform the specificity of the hybridization conditions and the alignment of the RNA size markers (data not shown). Membranes were also hybridized to rat glyceraldehyde 3-phosphate dehydrogenase (GAPDH) cDNA

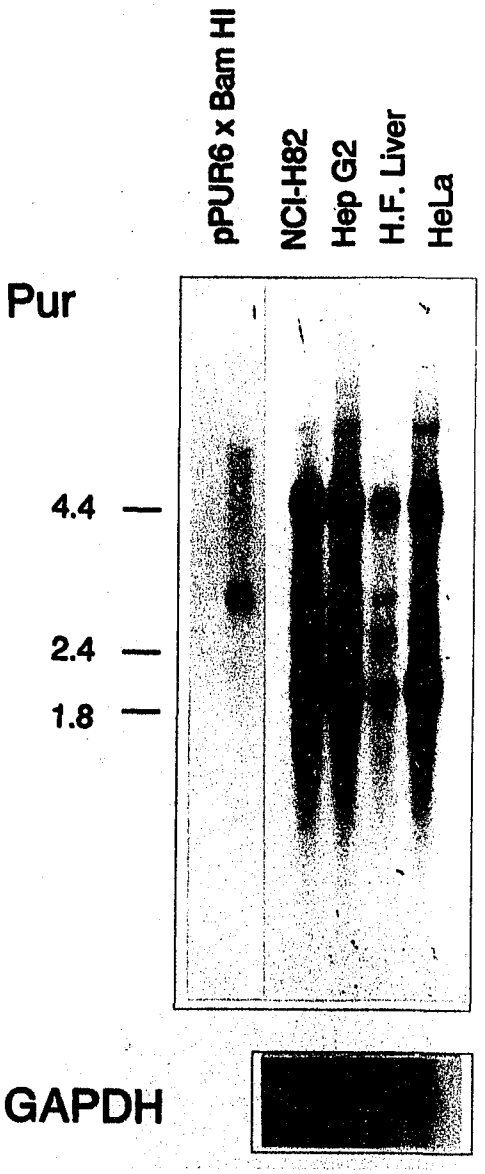

Fig 4. Northern analysis of human mRNAs with the Purα cDNA probe. Poly(A)$^+$ RNA prepared from tissue or culture cells was separated on 1.2% agarose gel containing 2.2% formaldehyde and blotted as described in Materials and Methods. Membranes were probed with labeled Purα cDNA. Lanes: pPUR6 x BamHI, pPUR6 DNA digested with

*Bam*HI; NCI-H82, lung tumor cell line mRNA; HepG2, hepatocacinoma 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 the left show molecular size markers in kilobases measured with the aligned 0.16-1.77 kb and the 0.24-9.5 kb RNA ladders (GIBCO-BRL).

(Fort et al., 1985) probe as a loading control (the lower box in Fig. 4). All the probes used here were labeled by random primer labeling. The overall intensity of bands from the fetal liver sample is lower than that from other cell types, and the intensity of the control GAPDH band for this sample is also lower. This may reflect the high level of liver-specific transcripts in the sample or the deviation of sample loading.

D. The size of the human Purα mRNA

Three human cDNA libraries (from HeLa cells, fetal liver cells, and testis cells) were screened using Purα cDNA (the *Pst*I fragment or the whole insert from λAB6) as the probe in order to get the missing cDNA sequence of Purα. All the positive plaques appeared to contain the same restriction map as λAB6 with similar or smaller size. This suggested that strong secondary structure be present in the Purα mRNA, which, during the process of establishing the cDNA libraries, prevent the elongation of the first strand cDNA synthesis over both the 3' and 5' regions near the Purα mRNA sequence correspondent to λAB6. As an alternative choice, rapid amplification of cDNA

ends (RACE) assays (Frohman et al., 1988) was performed to search for the possibly missing parts of Purα cDNA sequence in the 5' and 3' ends (see Fig. 1).

RACE assays are advantageous to obtaining full-length cDNA clones of a low-abundance mRNA when part of the cDNA sequence is known. RACE techniques employ 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. Synthesis of the first strand cDNA using an cDNA internal primer may significantly increase the target to background ratio. During the process of the first strand cDNA synthesis with a Purα cDNA internal primer, the following effects are expected to overcome the predicted secondary structure of the Purα mRNA: (1) annealing of the primer to the Purα mRNA at the location close to the sequence of secondary structure; (2) higher affinity between a DNA primer and a RNA homologous sequence; (3) high concentration of the Purα-specific primer used. Control experiment of RACE assays with the human α_1-antitrypsin (AAT) was successful, which indicated the practicability of this procedure under the conditions tested (data not shown).

3' RACE revealed three bands whose size was 420, 200, and 120 nucleotides, respectively (Fig. 5B). This may add at most 300 bp to the identified cDNA sequence of Purα at the 3' end. The multiple bands in the 3' RACE may represent different poly(A) addition sites or alternate splicing for the same gene, or result from short stretches of $(dA)_n$ in the 3'

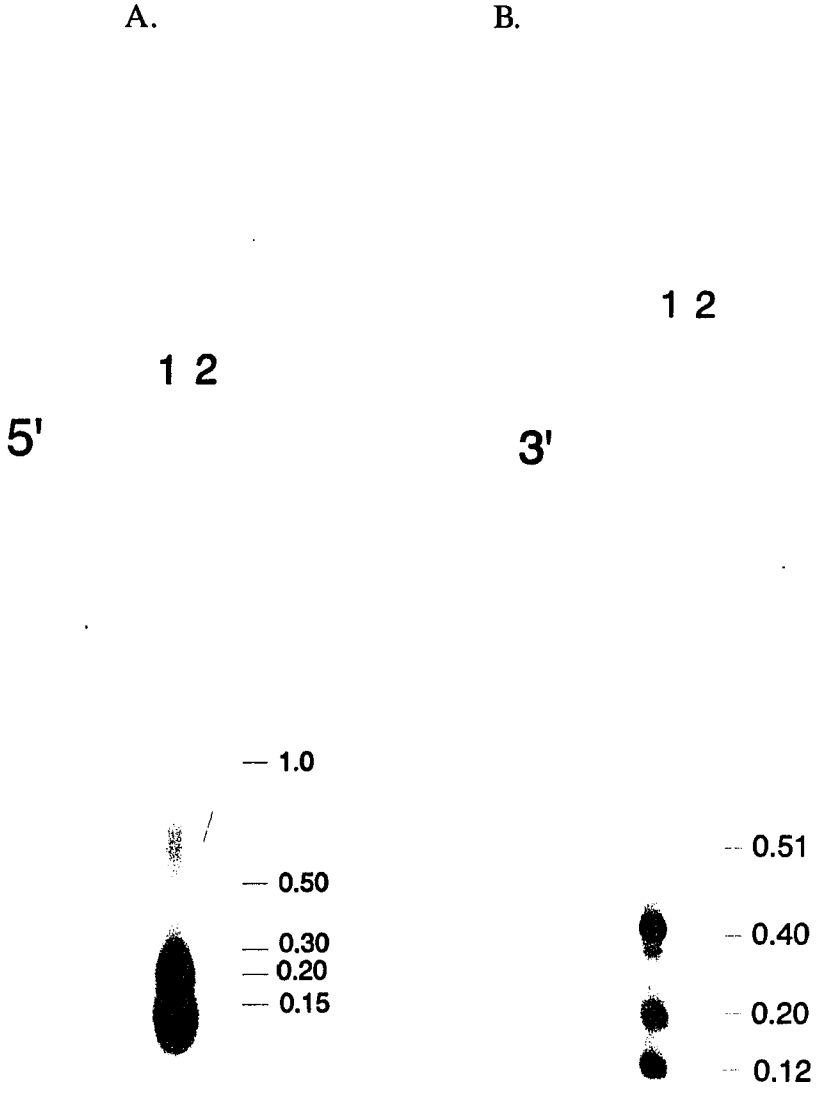

Fig. 5. 5' and 3' RACE extension of Purα cDNA. Products of the RACE reactions outlined in Materials and Methods were subjected to electrophoresis in 1.5% agarose gel, Southern blotted to GeneScreen Plus membranes, and hybridized to a Purα cDNA probe. In each case,

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 (refer to Table 1 for the sequences of primers used here) 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, corresponding to nucleotides 270 to 253, for the first amplification, and EX-174, corresponding to nucleotides 174 to 157 plus a *Hind*III linker, for the second amplification reaction. For lane 2, primer EX-174 was omitted. Right: 3' RACE. *pur*-specific primers used were EX-695, corresponding to nucleotides 695 to 714, for the first amplification, and EX-990, corresponding to nucleotides 990 to 1009, for the second amplification. For lane 2, primer PDT-01 was omitted.

cDNA sequence, or both. Further studies of genomic cloning of Purα gene and "exon trapping" (Buckler, et al., 1991) will distinguish among these possibilities. 5' RACE revealed two major bands at 50 and 250 nucleotides, and a diffuse range of much less intense bands at 500 to 700 nucleotides (Fig. 5A). Two smaller bands correspond with the 5' sequence shown in Figure 2 and could represent premature stopping of reverse transcriptase at the GC-rich sequence sites. The diffuse smear at 500 to 700 nucleotides could represent 5' end of Purα mRNA, and its lower intensity could explain the poor efficiency for reverse transcriptase to pass through secondary structure around the GC-rich region (upstream of the first translation codon, see Fig. 1). This may add at most 500 nucleotides to known cDNA sequence of Purα at the 5' end. These 3' and 5' RACE results may place the size of the Purα mRNA around 1.9 kb, which is consistent with results obtained by Northern blot analysis.

The 5' RACE products were purified from agarose gels (using DEAE NA45 membrane) according to their size, cloned into pBluescript II KS vector, and sequenced. No additional sequence was obtained.

2. Characterization of mouse Purα cDNA

A. Isolation of a Purα clone from a mouse λZAP II library by DNA-DNA hybridization

When human genomic Southern blot was performed, (for details please see Analysis of human chromosomal Pur gene) mouse genomic DNA was aligned in order to study the conservation of human Purα gene during evolution. There were 1 or 2 bands in each lane for restriction digestion of mouse genomic DNA (Fig. 8), which indicates that the mouse genome contains at least one copy of homologous sequence to the human Purα cDNA.

A mouse fetal heart cDNA library (prepared in λZAP II vector) was screened for mouse Purα cDNA clones. The screening conditions were the same as previously described (Bergemann, et al., 1992). The 777-bp *Pst*I fragment of the human Purα cDNA (nucleotides 165 through 941 in Fig. 2) was ^{32}P-labeled and used as the probe. A total of 3 x 10^5 recombinants were screened and one positive clone, LKDEL76, was isolated after three rounds of consecutive screening.

B. Sequence analysis of mouse Purα cDNA: Extraordinary degree of conservation between human and mouse Purα

The cDNA insert in LKDEL76 was subcloned into pBluescript II SK- phagemaid vector (this construct was referred to as PKDEL76) and sequenced. The nucleotide

```
-518                    CG CTA GGG AAG GAG GAG AGA GAG GGA GAG AGA
-486  GAG AGC TGA CGG GGG CTG TGA GTG AGC GGG AGG AGC GGG GTG TCA
-441  GCC CGG CCG GCC CGC GGG AGT GAA AGG TAG AGC GAA CGA GGC CCC
-396  GAG CCC GCC GCG CGG CCG CCG CCT CCG CCT CCT CCC CTC CCC TCG
-351  CGC GCG GCC CGC CCC CGC CCG CAG CCC GCC CTC CTC GCG GCT CAG
-306  CCC GCA GAG CCC GAG CGA GGC GCC GTT CGG CTC GCG CGC CGT CCA
-261  GAA CAG GTG TGC GCG CGC CTC CCC CCT CCC TCC CTC CTC CTC CTC
-216  GCG CCC GCC TGC GCG CGC CGG CCT CCC TCC AGC CTG CGC CCT CCC
-171  TCA CCT GCG GCA GGA CAG CGC CCG CCA GCC CGC CCG TCC CGA TCT
-126  CGC GAG AGT GGC TGA CTG GCT GTG GGG GTT GCG GCG GCA GCA GGC
 -81  GGA GCC GGG GAG GGA AAG CAG CGG CGG CTG AGG CGA CTG AGG CGG
                      . ... ... ... ... ... ... ... ... ... ...
                                                  <--

 -36  CGG GCG GAG CGG CAG ---------- GCG GCG GCG GAG CGC AGC ATC
      ... ... ... ... ... gcggcggcggc ... ..a ... ... ... ... ...

   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

        t                   g           c
  46  GGC TCG GGC GGC TCC CTA GGG CAC CCG 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

              c ggg       a
 133  GGC GGC GGG --- GCC CCG GGG GGG CTG CAG CAC GAG ACG CAG GAG
      Gly Gly Gly     Ala Pro Gly Gly Leu Gln His Glu Thr Gln Glu
                ***
                                                    c
 178  CTG GCC TCC AAG CGG GTG GAC ATC CAG AAC AAG CGT TTC TAC CTG
      Leu Ala Ser Lys Arg Val Asp Ile Gln Asn Lys Arg Phe Tyr Leu
```

```
                                c           c                   c
223  GAC GTG AAG CAG AAC GCT AAG GGC CGT TTC CTG AAG ATC GCA GAG
     Asp Val Lys Gln Asn Ala Lys Gly Arg Phe Leu Lys Ile Ala Glu

               g                                 t   t               a
268  GTG GGC GCT GGC GGC AAC AAG AGC CGC CTC ACC CTC TCC ATG TCT
     Val Gly Ala Gly Gly Asn Lys Ser Arg Leu Thr Leu Ser Met Ser

313  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

                                        g   g                   g
358  TAC GCG CAG CTG GGC CCC AGC CAG CCA CCC GAC CTG GCC CAG GCA
     Tyr Ala Gln Leu Gly Pro Ser Gln Pro Pro Asp Leu Ala Gln Ala

                   g                    a
403  CAG GAC GAG CCA CGC CGG GCG CTC AAG AGC GAG TTC CTG GTG CGC
     Gln Asp Glu Pro Arg Arg Ala Leu Lys Ser Glu Phe Leu Val Arg

       g
448  GAA 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

                                    g                   t
493  CGC TTC CTG CGC ATC CGC CAG ACA GTC AAC CGG GGG CCC GGC CTG
     Arg Phe Leu Arg Ile Arg Gln Thr Val Asn Arg Gly Pro Gly Leu

                                                      g       g
538  GGC TCC ACG CAG GGC CAG ACC ATT GCG CTG CCC GCA CAG GGT CTC
     Gly Ser Thr Gln Gly Gln Thr Ile Ala Leu Pro Ala Gln Gly Leu

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

628  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

                                                    c
673  GTG GAC AAC AAG CGC TTC TTC TTC GAT GTG GGT TCC AAC AAG TAC
     Val Asp Asn Lys Arg Phe Phe Phe Asp Val Gly Ser Asn Lys Tyr

                                  c                       t
718  GGC GTG TTT ATG CGA GTG AGT GAG GTG AAG CCC ACC TAC CGC AAC
     Gly Val Phe Met Arg Val Ser Glu Val Lys Pro Thr Tyr Arg Asn

                     c
763  TCC ATC ACC GTG CCC TAC AAG GTG TGG GCC AAG TTC GGA CAC ACC
     Ser Ile Thr Val Pro Tyr Lys Val Trp Ala Lys Phe Gly His Thr

                        g                                   g
808  TTC TGC AAG TAC TCC GAG GAG ATG AAG AAG ATT CAA GAG AAA CAG
     Phe Cys Lys Tyr Ser Glu Glu Met Lys Lys Ile Gln Glu Lys Gln

                        t   c               t
```

```
853   AGG GAG AAG CGG GCC GCT TGT GAG CAG CTC CAC CAG CAG CAA CAG
      Arg Glu Lys Arg Ala Ala Cys Glu Gln Leu His Gln Gln Gln Gln

                g               g               t                   t
898   CAG CAG CAA GAG GAG ACC ACC GCT GCC ACC CTG CTA CTG CAG GGC
      Gln Gln Gln Glu Glu Thr Thr Ala Ala Thr Leu Leu Leu Gln Gly
                              ***

943   GAG GAA GAA GGG GAA GAA GAT TGA TCA AAC TGA ATG AAA CAC ACA
      Glu Glu Glu Gly Glu Glu Asp End
                                      ... ... a.. ... ... .c. c..

988   CAC ACA CAC ACA CAC GCA TAC ACA TAC GTG TAC ACA CAC ACA CAC
      ... ... ... ... tg. at. c.. ... c.. aca c.g c.. ... ... ..-

1033  ACA GCC ACA CAC AGA GAA AAT ATA CTG TAA AGA GAG AAA ATA AAA
      --- --- --- --- --- ... ... ... ... ... ... a.. .g. ga. .t.

1078  AGT TAA AGG AAT TC
      .aa agt taa ..a gtt aaa aaa aaa aaa a
                             -->
```

Fig. 6. Nucleotide sequence of mouse Purα and comparison between the mouse and human *purα* gene. Nucleotide sequence of mouse Purα are presented in upper case letters. The corresponding human cDNA sequence, where known, is presented in lower case letters. Nucleotide identities in the non-coding regions are shown by dotted lines. Gaps in either the mouse or human sequence are shown by dashed lines. In the 5' leader sequence in-frame stop codons are underlined. In the coding region nucleotides which differ between the mouse and human genes are underlined, and the human version is presented above in lower case. Amino acids which differ between mouse and human are starred. The heads of dashed arrows indicate the beginning points of new human cDNA sequence that was not originally published. Both mouse and human Purα cDNA sequences have been deposited with GenBank (accession #s U02098 and M96684, respectively).

sequence of the cDNA insert in PKDEL76 was determined and deposited with the GenBank (GenBank Accession #U02098).

The cDNA is 1609 bp in size, including 5' leader sequence of 519 bp and a 3' untranslated sequence of 125 bp. The 5' leader sequence contains stop codons in all three reading frames. This suggests that the open reading frame of Purα is

complete. The 3' untranslated sequence contains a polyadenylation signal. The protein coding sequence (321 aa) has been compared to that of the human gene (322) revealing extraordinary conservation (99.4%) of the protein (Fig.6). There are only two amino acid differences between the mouse and human Purα proteins. Human and mouse Purα cDNAs are over 96.3% identical within the coding sequence. There are 36 single-nucleotide changes between the two coding sequence, but only one of these, G to A at nucleotide 916, results in alteration of an amino acid, alanine to threonine at codon 306 of the mouse gene. Additionally, the human gene contains an extra glycine codon (GGG) within a multi-glycine repeat extending from mouse codon 44 through 48.

C. Conservation in human and mouse of a motif common to several proteins involved in initiation of DNA replication

The mouse and human Purα proteins and the human Purβ protein possess a region of limited but significant homology to simian virus 40 (SV40) large T-antigen, the viral replication-initiation and cell-transformation protein, as revealed by the Blast algorithm (Altschul et al., 1990). The probability of such a match between these two proteins is approximately 10^{-5}. Purα and SV40 T-antigen both contain a sequence of precisely 22 amino acids beginning with PTY and ending with FC in which there are 5 identities between the two

proteins and several conservative substitutions. Allowing for a two-amino-acid gap, there are 9 identities within 28 amino acids. A comparision of this motif, present in transforming proteins of several DNA tumor viruses, as well as in proteins of yeast and human cells, is presented in Fig. 7, and a consensus sequence has been derived. Since this consensus contains PSY and C, it is termed the "psycho" motif. Analysis of T-antigen has implicated that at least a portion of this motif is involved in protein-protein interactions since LFCSE is important for binding to the retinoblastoma protein pRB (DeCaprio et al., 1988; Wang et al., 1993). The Pur proteins do not possess precisely this configuration although FC and SE are present near the carboxyl terminus of the psycho motif. The 4 C-terminal amino acids of this region, SEEM, are identical for Purα and SV40 T-antigen. While the serine in this motif is a known site of casein-kinase II-catalyzed phosphorylation in T-antigen (Graesser et al., 1988), it is not yet known whether this serine is phosphorylated in Purα.

Although the psycho motifs of the E7 proteins of human papilloma virus type 58 and Rhesus monkey papilloma virus are the least significantly homologous among the group presented, the E7 proteins still possess certain structural features in common with all of the listed proteins. For example, the YEQL configuration exactly matches the consensus for all of the proteins. The distribution of aromatic and hydrophobic amino acids is also similar as described below. The E7 proteins of 25 papilloma virus subtypes are presently detected by

BLAST in various databases. While some of these are more homologous to the Pur proteins, the two presented are representative of both homologies and variability among E7 proteins. All of the E7 proteins possess the configuration LXCXE, deemed important for pRB binding, near the C-terminus of the psycho motif.

Proteins Erg1 and Erg2 are members of the *ets* gene family (Rao et al., 1987). In these proteins, which are transcribed from the same gene, the motif is part of a region of homology to the v-*ets* gene which also has transcription activation properties (Siddique et al., 1993). The Erg psycho motif is included in a helix-loop-helix structure in which PSY is part of a loop configuration preceding an alpha helix (Schneikert et al., 1992). The Erg proteins do not possess the configuration LXCXE, but rather a similar one of LTSDD. At least one *ets* protein, Elf-1, does possess LXCXE in a region of *ets* homology, and it is known to bind pRB (Wang et al., 1993).

It is unlikely that the psycho motif functions as an essential pRB-binding motif since several pRB-binding proteins do not have a strong version of the motif. It is conceivable, however, that the motif functions as modulator of pRB binding in pRB-binding proteins which do possess the configuration LXCXE, and all bind pRB. However, there are differences in the affinities of E7 proteins from different HPV subtypes for pRB, and these could be correlated with differences in psycho motifs among these subtypes. A number of proteins which bind pRB, including HPV E7, contain amino acid residues in common in

addition to LXCXE, and several of these are in the psycho motif regions (Wang et al., 1993).

The PSY and FC groups, near the beginning and end of the psycho motif, are most strongly conserved. (Of all proteins shown only the yeast protein MCM2 does not possess the initial P, and it is evolutionarily distant from the mammalian proteins.) It is likely that a specific spatial relationship between these groups must be preserved. In the case of the papovavirus T-antigens, the Erg proteins and the Pur proteins, a loop and helix structure could serve this purpose, while for the E7 proteins, looping caused by multiple proline residues could achieve the same effect.

Every one of the 14 sequences listed in Fig. 7 contains at least 3 aromatic or bulky hydrophobic aa spaced 3 or 4 residues apart. Alpha helicity is predicted in this region for Erg1 and 2, the T-antigens and Purα and Purβ (Chou and Fasman, 1974; Rost and Sander, 1993). This spacing could thus present aromatic residues along one face of a helix. It has previously been predicted that this region forms an amphipathic alpha helix in Purα and Purβ with aromatic aa on one face of the helix and basic aa on another (Bergemann et al., 1992). In the E7 proteins the potential alpha helix is confined to the amino terminal region of the psycho motif, proline residues interrupting the carboxyl terminus, and the helix contains three hydrophobic aa spaced as described. In Purα potential uninterrupted helix extends through the SEEM configuration and includes six aromatic or hydrophobic aa

spaced as described. Four of the human cellular proteins possess two aa between C and the SEEM homology. In each case these residues extend the potential helix with aromatic or hydrophobic residues spaced as described. These additions also intervene in the LXCXE configuration thought to be necessary for pRB binding. It will be interesting to see whether any of these proteins binds pRB, especially since several of these proteins may be involved in initiation of DNA replication.

The MCM proteins of yeast belong to an extensive class of nucleic acid-dependent ATPases involved in the initiation of DNA replication in both prokaryotes and eukaryotes (Koonin, 1993). MCM2 is reported to be important for *ARS*-dependent initiation in yeast (Yang et al., 1991). Purα, which binds an element present at several recently-mapped mammalian origins of replication, has regions of homology to the ATP-binding motif of MCM proteins and of other proteins involved in initiation of replication, including the DnaA protein of *E. coli* (Koonin, 1993). BM28, a human MCM2 homolog, is reportedly required for entry of cell into S-phase (Todorov et al., 1994). It is interesting in this regard that these cellular proteins possess homologies to SV40 large T-antigen, a well-characterized replication initiation protein and transcription factor. The viral transforming proteins in Fig. 7 do not all affect initiation of replication directly. The E7 proteins are not known to affect initiation through binding to an origin of replication, although they may exert an indirect effect via their interaction with pRB. Similarly, the Erg proteins, known transcription factors, may

influence DNA replication indirectly via effects on the transcription of certain genes. Studies of Purα deletion mutants indicate that the psycho motif is not required for sequence-specific DNA binding by this protein. The comparison in Fig. 7 points toward investigation of a role for this motif in protein-protein interactions that could modulate either transcription or the initiation of DNA replication or both.

SV40 T-Ag	84	PTYGTDEWEQWWNAFN---EEN-LFC--SEEM	109
JCV	83	PTYGTDEWESWWNTFNEKWDE-DLFC--HEEM	111
BKV	83	PTYGTEEWESWWSSFNEKWDE-DLFC--HEDM	111
Bov. polyoma	70	PEYGTSQWEQWWEEFNQGFDEQDLHC--DEEL	99
Mus. polyoma	83	PSYGTPEWDEWWKEFNKDF---DLFC--NEAF	109
Monk. Lympho.	108	PNPGEGSWGKWWREFVNRQCCDDLFC--SETM	137
Rh. PV E7	6	PTLEDIVLD--LQPF-PQPQPVDLMC--YEQL	32
HPV-58 E7	16	PTLTEYILD--LH-----PEPTDLFC--YEQL	38
Erg1	89	PSYNADILLSHLH-Y-LRETPLPHLT--SDDV	116
MCM2, yeast	659	STLPLAQNVSLTEPILSRF---DILCVVRDLV	687
ORF, myel.	649	PSLTFSENVDLTEPIISRF---DILCVVRDTV	677
BM28	628	PSLTFSENVDLTE-IISKF---DILCYDRDTV	656
Purβ	54	PSYRNAITVPF-KAW-GKFGGT--FCRYADEM	81
Purα	251	PTYRNSITVPY-KVW-AKFGHT--FCKYSEEM	278

0.4	PTYGT·EWe·WWK·FN·KFe··DLFC···EEM
0.5	PTY····&e·ϕ&··F··KF···DLFC ·EE&
0.6	PST····&··ϕ&··ϕ··Kϕ···DLFC ·E·&
0.7	PS&····&··&&··ϕ ····· D&&C ·e·&
0.8	PS&·····&···& ···· ·&&C ·e·&

Fig. 7. The psycho motif is common to Purα, Purβ, transforming proteins of DNA tumor viruses and cellular proteins regulating

transcription and/or replication. Abbreviations at left in Fig. 7 are for the following proteins. SV40 T-Ag: the Rb-binding motif of SV40 large T-antigen; JCV: polyoma virus JC T-antigen; BKV: polyoma virus BK T-antigen; Bov. polyoma: bovine polyoma virus T-antigen; Mus. polyoma: mouse polyoma virus T-antigen; Monk lympho: monkey B-lymphotropic papovavirus T-antigen; Rh. PV E7: Rhesus monkey papilloma virus protein E7; HPV-58 E7: human papilloma virus type 58 protein E7; Erg 1: human *ets* proteins Erg1; MCM2, yeast: the MCM2 protein of S. cerevisiae; ORF myel.: an open reading frame for an as yet unknown protein form human myeloblasts; BM28: the human BM28 protein; Purβ: human protein Purβ; Purα: human and mouse proteins Purα. Codon #'s for each protein sequence are given at the immediate left and right of the sequence. Codon #'s for Purβ are from the first available aa since the complete amino terminus has not been sequenced. Database accession numbers are as follows: SV40 T-Ag, gb:V01380; JCV, sp:P03072; BKV, pir:A33278; Bov. polyoma, gb:S48202; Mus. polyoma, gb:M55904; Monk. lympho., gb:K02562; Rh. PV E7, pir:B38503; HPV-58 E7, gb:D90400; Erg1,2, gb:M21535; MCM2,yeast, gb:X53539; ORF, myel., gb:D21063; BM28, embl:X67334; Purβ, gb:B45036; Purα, gb:M96684. Abbreviations for the databases are: gb, GenBank; pir, PIR; embl, EMBL. Sequences have been aligned to maximized the number of overall homologies. Sequence consensus is analyzed below. Percentages of all analyzed sequences are given at left. Aa present in > 90% of analyzed sequences are underlined. A dot at a position indicates any aa residue. The following notation, adapted in part from Koonin (1993), is used: s = S or T; e = E or D; k = K or H; & indicates a bulky hydrophobic residue, aliphatic or aromatic (I, L, V, M, F, Y, W); ϕ = an aromatic residue (F,Y,W).

3. Analysis of the human chromosomal *pur* gene

A. Southern analysis of human genomic DNA: human genome contains no more than two different copies of homologous sequences to the Purα cDNA/*Pst*I fragment

Northern analysis of Pur has revealed multiple transcripts with similar pattern for each cell line and tissue studied (Fig. 3), which has raised the following question: do

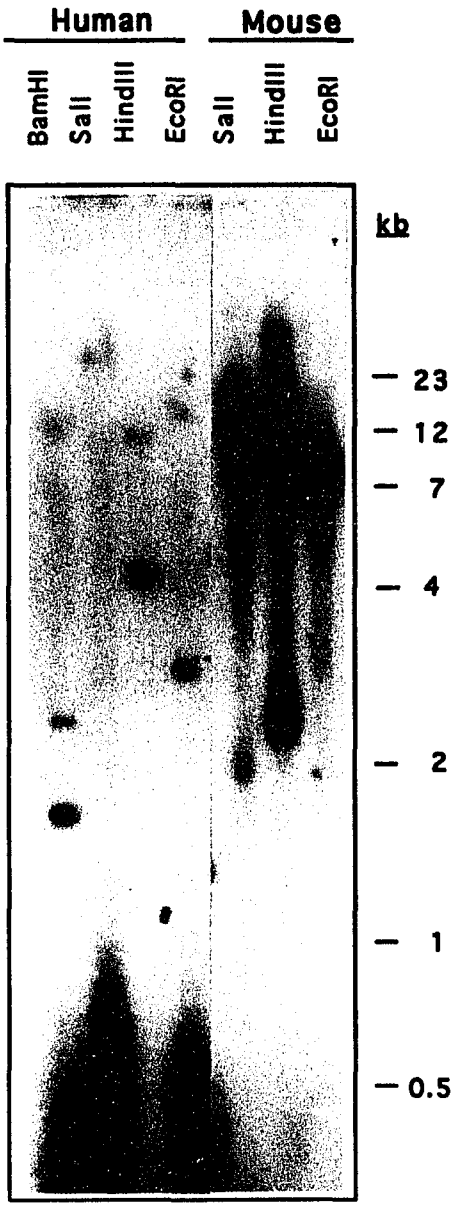

Fig. 8. Southern analysis of human and mouse genomic DNA with single restriction endonuclease digestion. Ten µg of genomic DNA from HeLa cells or 10 µg of genomic DNA from mouse C2 myoblast cells was digested with individual restriction endonuclease (*BamH*I, *Sac*I, *Hind*III, and *E c o R* I) as indicated at the top of the panel. The

restriction fragments were separated on 0.8% agarose gel and blotted to GeneScreen Plus Membrane. Membranes were probed with labeled 777-bp Purα cDNA fragment as described in Materials and Methods. Numbers on the right show DNA molecular size markers in kilobases measured with the aligned 1 kb DNA ladder and λ DNA/*Hind*III fragments (GIBCO-BRL).

the multiple Pur related transcripts result from alternate processing of the same pre-mRNA from a single gene, or represent a family of *pur* related genes, or both? To address this question, Southern analysis of HeLa cell genomic DNA with single or double restriction endonuclease digestion was performed as a preliminary study (Fig. 8 and Fig. 9). The probe used was the labeled 777-bp *Pst*I fragment of the human Purα cDNA.

The lane of *Hind*III digest revealed 2 bands (Fig. 8), 11.5 kb and 4.5 kb. The lane of *Pst*I digest revealed 2 bands (Fig. 9), 4.0 kb and 0.8 kb. The lane of *Hind*III + *Pst*I digest revealed 2 bands(Fig. 9), 1.6 kb and 0.8 kb. These results suggest that human genome contains no more than 2 different copies of homologous sequence to the Purα cDNA.

There were 1 or 2 bands for each restriction digestion of mouse genomic DNA (Fig. 8). *Sal*I digest revealed one band of about 15 kb in size. *Hind*III digest revealed two bands, approximately 15 kb and 2.5 kb, respectively. *EcoR*I digest revealed one band of approximate 8 kb. This result indicates that mouse genome contains at least one, and no more than two, copy of homologous sequence to the Purα cDNA/*Pst*I fragment.

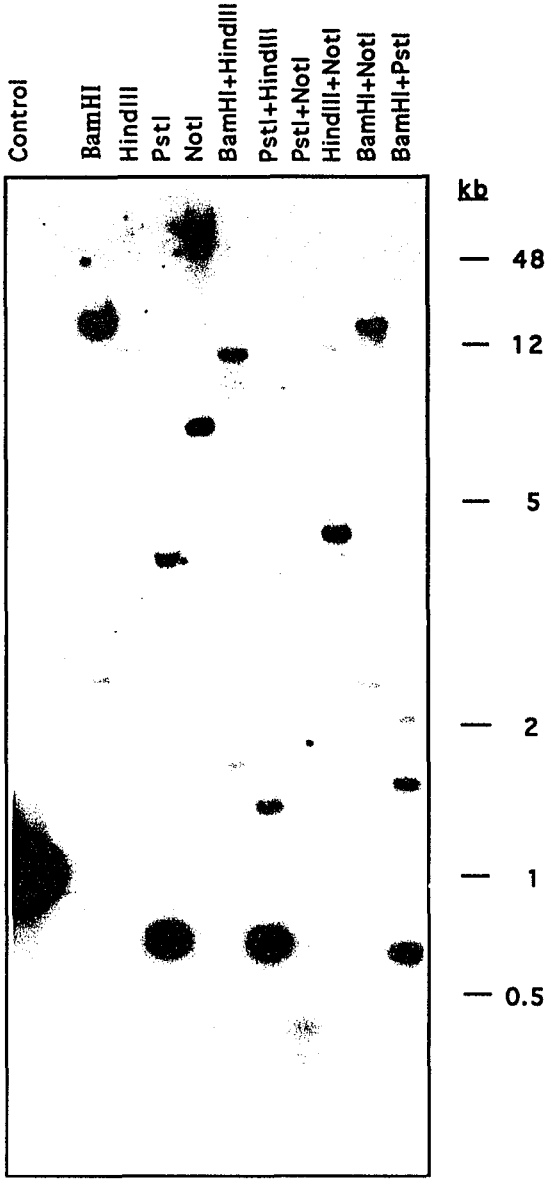

Fig. 9. Southern analysis of human genomic DNA with single and double restriction endonuclease digestion. Ten μg of genomic DNA from HeLa cells was digested with restriction endonuclease(s) as indicated at the top of the panel. Control line was 10 pg of human Puroα cDNA insert.

The restriction fragments were electrophoresed on 0.8% agarose gel, transfered to the membrane, and probed with labeled 777-bp Purα cDNA fragment at the same conditions as that in Fig. 8. Numbers on the right show DNA molecular size markers in kilobases measured with the aligned 1 kb DNA ladder and λ DNA/*Hind*III fragments (GIBCO-BRL).

B. Genomic cloning of human Pur gene

Southern analysis of human gemonic DNA alone is limited in revealing the copy number of homologous sequence to the *purα* gene. A human placenta genomic library (prepared in EMBL3 vector, Clontech) was screened for genomic clones of *pur*. This library was generated by cloning *Sau3A*I partial digests of human genomic DNA into the *BamH*I sites of the EMBL3 vectors (Frischauf et al., 1983). There are two *Sal*I sites located external to each of the *BamH*I site in the polycloning-site sequence of the vector. Cloned fragments can be excised from recombinants by digestion with *Sal*I. A total of 1.6 x 10^6 plaques were screened with the labeled 777-bp *Pst* I fragment of the human Purα cDNA. Seven positive clones (designated LGT1, LGT2, LGT3, LGT6, LGT7, LGT9, and LGT12) were isolated after 3 rounds of consecutive screening. These recombinant phage DNAs were individually digested with *Hind*III, *Pst*I, and *Hind*III + *Pst*I. Southern blots of those restriction digests (with the 777-bp *Pst*I fragment as the probe) revealed that four clones (LGT1, LGT2, LGT3, and LGT12) each contained the 0.8 kb *Pst*I fragment and a *Hind*III fragment larger than 12 kb (instead of 11.5 kb) with variable length (the EMBL3 vector

may contribute to one of the *Hind*III sites; if this is the case, the *Hind*III fragments all contain a 4.42 kb vector sequence); and other three clones (LGT6, LGT7, and LGT9) each contained the 4.5 kb *Hind*III fragment, 4.0 kb *Pst*I fragment, and 1.6 kb *Hind*III + *Pst*I fragment (Fig. 10). These results were consistent with that of Southern analysis of human genomic DNA shown in Fig. 8 and Fig. 9.

In addition, when the same blot was washed off the probe and re-hybridized to the labeled 0.17 kb fragment 5' to the first *Pst*I site within the Purα cDNA sequence (Fig. 1), only LGT1, LGT2, LGT3, and LGT12 hybridized to the probe, while none of LGT6, LGT7, or LGT9 hybridized to the probe (Fig. 11). This result indicates that the 5' Purα cDNA sequence is much less, or not, homologous to the Purβ cDNA sequence, in contrast to the *Pst*I fragment. However, this result also may indicates that the 5' *pur*β sequence is not located within the range of those genomic clones.

These data allowed me to correlate the various restriction patterns of the *pur* gene in human genomic DNA with that of the human genomic clones. When these results are compared with those in Fig. 8 and Fig. 9, it can be deduced that the 11.5 kb *Hind*III fragment, the 0.8 kb *Pst*I fragment, and the 0.8 kb *Hind*III + *Pst*I fragment stay closely together in one genomic fragment, while the 4.5 kb *Hind*III fragment, the 4.0 kb *Pst*I fragment, and the 1.6 kb *Hind*III + *Pst*I fragment stay closely together in another genomic fragment. This could mean that they are two genomic fragments belonging to a single gene (the

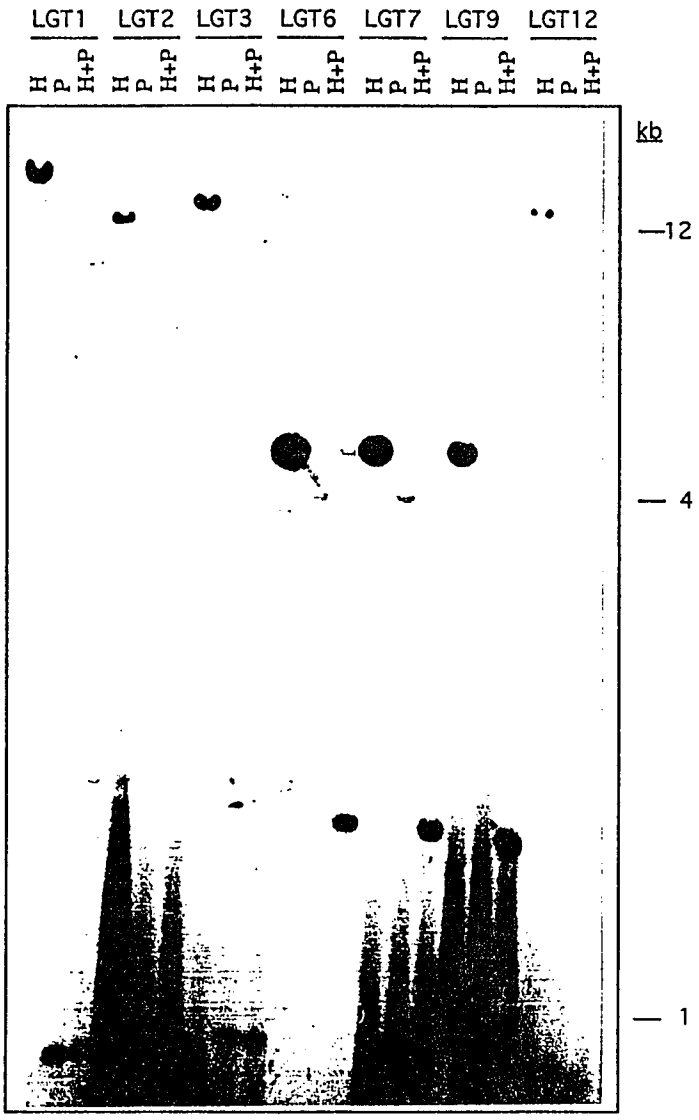

Fig. 10. Southern analysis of human genomic clones with the 777-bp *Pst*I fragment as the probe. Ten μg of recombinant λ phage DNA was digested with restriction endonuclease as indicated at the top of the panel. Abbreviations at the top panel are for the following restriction endonucleases. H: *Hind*III; P: *Pst*I; H + P: *Hind*III and *Pst*I. The restriction fragments were electrophoresed on 0.8% agarose gel, transfered to the membrane, and probed with the labeled 777-bp Purα cDNA fragment at the same conditions as that in Fig. 8. Numbers on the right show molecular size markers in kilobases measured with the aligned 1 kb DNA ladder and λ DNA/*Hind*III fragments (GIBCO-BRL).

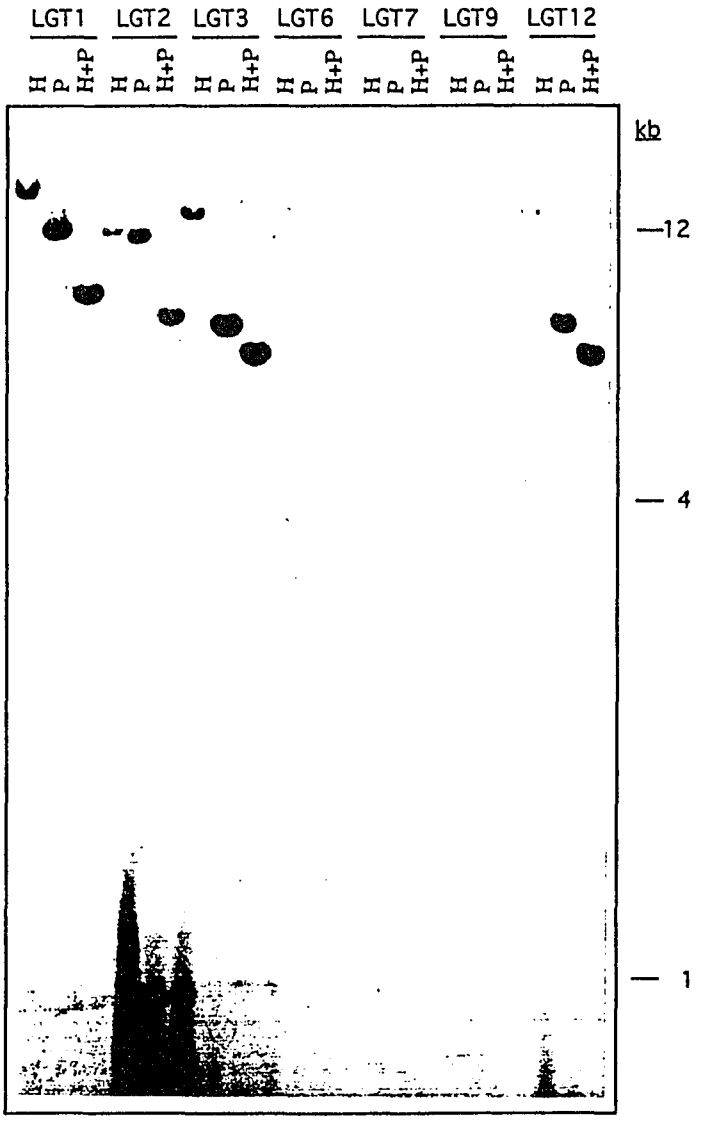

Fig. 11. Southern analysis of human genomic clones probed with the 0.17 kb Puro cDNA 5' to the *Pst*I fragment, as illustrated in Fig.1. The same membrane shown in Fig.10 was washed off the probe, and re-hybridized to the 5' Puro cDNA probe at the same conditions. Abbreviations at the top panel are for the following restriction endonucleases. H: *Hind*III; P: *Pst*I; H + P: *Hind*III and *Pst*I. Numbers on the right show molecular size markers in kilobases measured with the aligned 1 kb DNA ladder and λ DNA/*Hind*III fragments.

*pur*α gene) or that they are two genomic fragments belonging to two homologous genes (one of which is the *pur*α gene). To distinguish these two possibilities, partial sequencing of these genomic clones is required.

C. Partial sequencing of human genomic clones: both *pur*α and *pur*β genes were hybridized to the Purα cDNA/*Pst*I fragment

The 4.5 kb *Hind*III fragment from LGT6 was isolated, subcloned into pBluescript II KS vector (this construct was named PGT6H4.5), and partially sequenced. Sequence comparison between PGT6H4.5 and the human Purβ cDNA sequence (406 bp of human Purβ cDNA sequence is known, Bergemann et al., 1992) revealed that PGT6H4.5 contained the correspondent Purβ cDNA sequence known from nucleotide 1 through 277. The beginning site is one of the *Hind*III cloning site which is located within the cDNA sequence. Digestion of PGT6H4.5 with *Hind*III + *Pst*I generated a 1.6 kb fragment, which was isolated and subcloned into pBluescript II KS vector (this construct was named PGT6PH1.6). PGT6PH1.6 was partially sequenced. Sequence comparison revealed that PGT6PH1.6 also contained the Purβ cDNA sequence, beginning with the *Hind*III cloning site, from known nucleotide 1 through 277. Those results confirmed that the 4.5 kb *Hind*III fragment, the 4.0 kb *Pst*I fragment, and the 1.6 kb *Hind*III/*Pst*I fragment, shown on the genomic Southern blots in Fig. 8 and

Fig. 9, were from the chromosomal gene of Purβ. Those results also confirmed that the LGT6, LGT7, and LGT9 contained the human *purβ* gene.

In order to further study LGT1 (as well as LGT2, LGT3, and LGT12), PCR was performed using LGT1 recombinant DNA as the template and 2 primers (NOTF and NOTR) from selected human Purα cDNA sequence. The recombinant plasmid DNA PZT6 (containing 1.12 kb of human Purα cDNA insert) was used as the template for the PCR control (lane RP16). The PCR products were electrophoresed on 1.2% agarose gel stained with ethidium bromide and photographed (Fig. 12). There was a 0.35 kb fragment for both sample lane RL15 and control lane RP16. There were three additional bands in the lane RP16, and all these bands appeared to be over the size of 2 kb when compared with the aligned DNA molecular size marker. Those three bands were believed to be the plasmid DNA PZT6 with different forms. The 0.35 kb PCR fragment was individually purified from agarose gel using the GeneClean method, and was directly sequenced using NOTF as the primer. The 310 nucleotide sequence obtained was identical to the Purα cDNA sequence from nucleotide 515 through 824 (including 14 nucleotide sequence from the primer NOTR). This confirmed that LGT1 contained the chromosomal *purα* gene, and suggested that the 0.35 kb cDNA sequence between primers NOTF and NOTR was located within a single exon of the *purα* gene. This also confirmed that the 11.5 kb *Hind*III fragment

and the 0.8 kb *Pst*I fragment shown on the genomic Southern blots in Fig. 8 and Fig. 9 were from the *pur*α gene.

Southern analyses of human genomic DNA and human genomic clones with partial sequencing of the genomic clones, as described above, confirm that *pur*α gene and *pur*β gene are the only two genes hybridized to the Purα cDNA probe. And

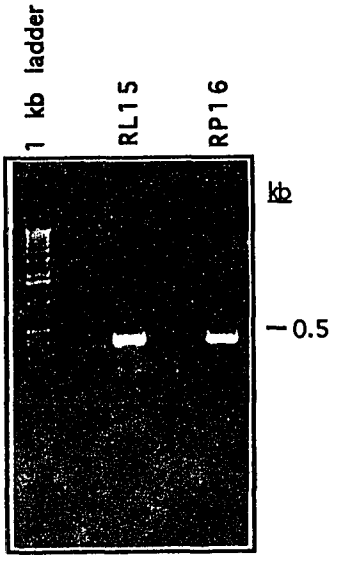

Fig. 12. Electrophoresis of PCR products on agarose gel. The PCR template is the recombinant phage DNA LGT1 for lane RL 15 and the recombinant plasmid DNA PZT 6 for control lane RP 16. Two primers from selected human Purα cDNA sequence were NOTF (GAACCAGCGCGGCCGCTTCCTGCG), corresponding to nucleotides 483 to 506, and NOTR (CATCTCCTTCGAGTACTTGCAGAA), corresponding to nucleotides 834 to 811. The PCR profile was as following: for 25 cycles, denaturation at 98°C for 1 minute, annealing at 60°C for 2 minutes and extension at 72°C for 2 minutes. Ten µl of PCR products was directly electrophoresed on 1.2% agarose gel, visualized after ethidium bromide staining. The DNA molecular size marker shown on the left is 1 kb DNA ladder. The number on the right shows the position of a 0.5 kb DNA fragment.

this may also suggest that only *purα* and *purβ* gene transcripts should be hybridized to the Purα cDNA probe as seen on the Northern analysis (Fig. 4).

D. Partial restriction maps of the human *purα* and *purβ* genes

Partial restriction maps of human *purα* and *purβ* genes, presented in Fig. 13 and Fig. 14, are obtained by summing up the information from Southern analysis of human genomic DNA and human genomic clones, and from partial sequencing of human genomic clones as described above. The orientation of these maps are determined based on that of correspondent cDNAs. In both Fig. 13 and Fig 14, the border of the exons is not determined yet.

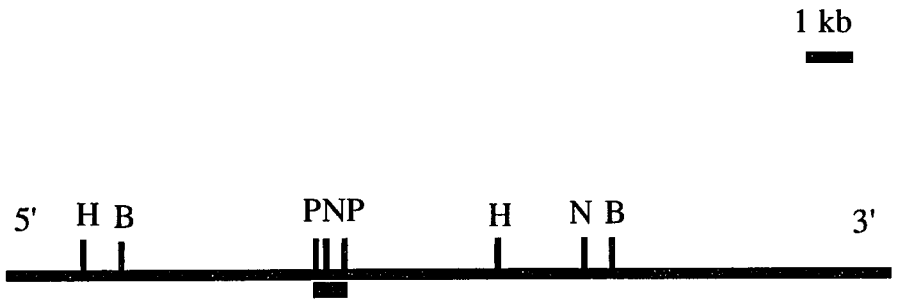

Fig. 13. Partial restriction map of human *purα* gene. Abbreviations are for the following restriction endonucleases. H: *Hind*III; B: *BamH*I; P: *Pst*I; N: *Not*I. The solid box at the bottom indicates the position of the known *purα* exon.

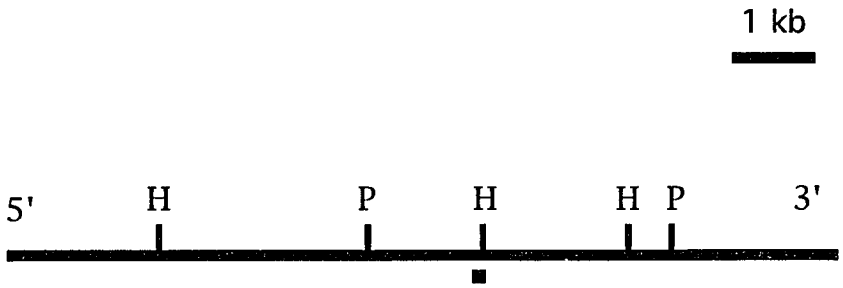

Fig. 14. Partial restriction map of human *purβ* gene. Abbreviations are for the following restriction endonucleases. H: *Hind*III; P: *Pst*I. The solid box at the bottom indicates the position of the known *purβ* exon.

E. Fluorescence *in situ* hybridization of the human chromosomal *purα* gene

In order to map the chromosomal location of the human *pur* gene, fluorescence *in situ* hybridization of a genomic DNA probe on human chromosomes was performed using LGT1, the recombinant EMBL3 DNA, as the probe. LGT1 was labeled with biotin-11-dUTP by nick-translation as described in Materials and Methods. Preparations of metephase-chromosome spreads, chromosomal *in situ* hybridization, and digital microscopic image analysis were performed as previously described (Najfeld et al., 1992). Chromosome identification and locus assignment was generated by hybridizing chromosomes with Alu sequence which was labeled with digoxigenin-11. The Alu sequence probe was visualized by adding a rhodamine-conjugated anti-digoxigenin antibody to produce a R-like

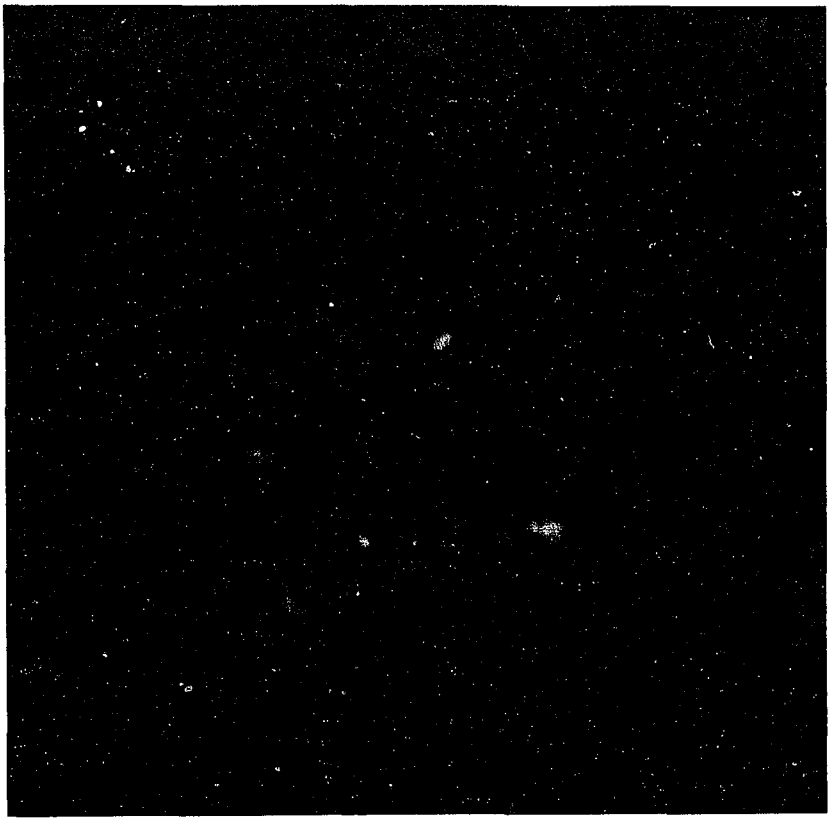

Fig. 15. Double-labeled fluorescence image of metaphase chromosome spread. DAPI G-like banding (blue) was applied to the metaphase chromosome spread. FITC fluorescence was pseudocolored red for best photographic contrast. The *pur*α genomic clone LGT1 was used as the probe as described in Materials and Methods.

Fig. 16. Localization of the whole genomic clone LGT1 to human chromosome 5. Alu sequence R-like banding (green/yellow) and DAPI G-like banding (blue) applied to two nonhomologous chromosome 5s in which the *pur*α locus (red) is visualized by *in situ* hybridization using *pur*α genomic clone LGT1 as the probe.

banding pattern. Chromosome identification was conformed by simultaneous diamidinophenylinodole (DAPI) staining, which produced a G-like banding pattern. Biotin-labeled human *purα* genomic DNA probe was visualized by using fluorescein isothiocyanate (FITC)-conjugated avidin DCS.

Peripheral blood cells from a healthy individual were stimulated with phytohemagglutinin (PHA) for preparation of metaphase chromosome spreads. Fifty-two metaphase chromosome spreads were randomly chosen and analyzed. Hybridization signals were detected on the long arm of chromosome 5 at band 5q31 in 14 metaphases, on the long arm of chromosome 6 at band 6q14 in 18 metaphases, and on both 5q31 and 6q14 in 20 metaphases.

In Fig. 15, one human metaphase chromosome spread is shown. The chromosome identities are shown as blue by DAPI banding. The chromosomal loci hybridized to the *purα* genomic DNA probe are shown as red dots (FITC-conjugated avidin for the biotin labeled DNA probe). There are four pairs of chromatids containing the signal of red dots. Two pairs of the red dots are located on the long arm of chromosome 5 at band 5q31 (upper left and lower right of the four pairs of chromatids). Another two pairs of red dots are located on the long arm of chromosome 6 at band 6q14 (upper right and lower left of the four pairs of chromatids).

In Fig. 16, two pairs of human nonhomologous chromosome 5s are shown with two sets of fluorescence image. The chromosomes on the left in each set are shown as

green/yellow (Alu sequence R-like banding) for the chromosome identities. The chromosomes on the right in each set are shown as blue (DAPI G-like banding) for the chromosome identities. The chromosomal loci hybridizing to the *purα* genomic DNA probe are shown as red dots (FITC-conjugated avidin for the biotin labeled DNA probe). The red dots are located on each pair of chromosomes at the band 5q31. A similar observation is obtained at band 6q14 at the same conditions (data not shown). Thus, the human *purα* gene is preliminarily mapped to 5q31 and 6q14 (Fig. 15 and Fig. 16).

F. Genomic Southern analysis with human/Chinese hamster somatic hybrid cells: Localization of human *purα* gene to chromosome 5

Genomic Southern analyses with genomic DNA of human/Chinese hamster somatic hybrid cells, which contain a single human chromosome 5 or 6, were performed to further study the chromosomal localization of human *pur* gene (Ma et al., 1994b). The genomic DNA was individually digested with either *Hind*III or *EcoR*I, and hybridized with the 777-bp *Pst*I fragment from the human Purα cDNA (Fig. 17). The 11.5 kb *Hind*III band, which was from the human *purα* gene, was present in the human DNA control (the lanes for HeLa and human placental DNA) and hybrid cell/chromosome 5, but absent in hamster DNA control (CHO lane) and hybrid cell/chromosome 6. These results indicate that the human *purα*

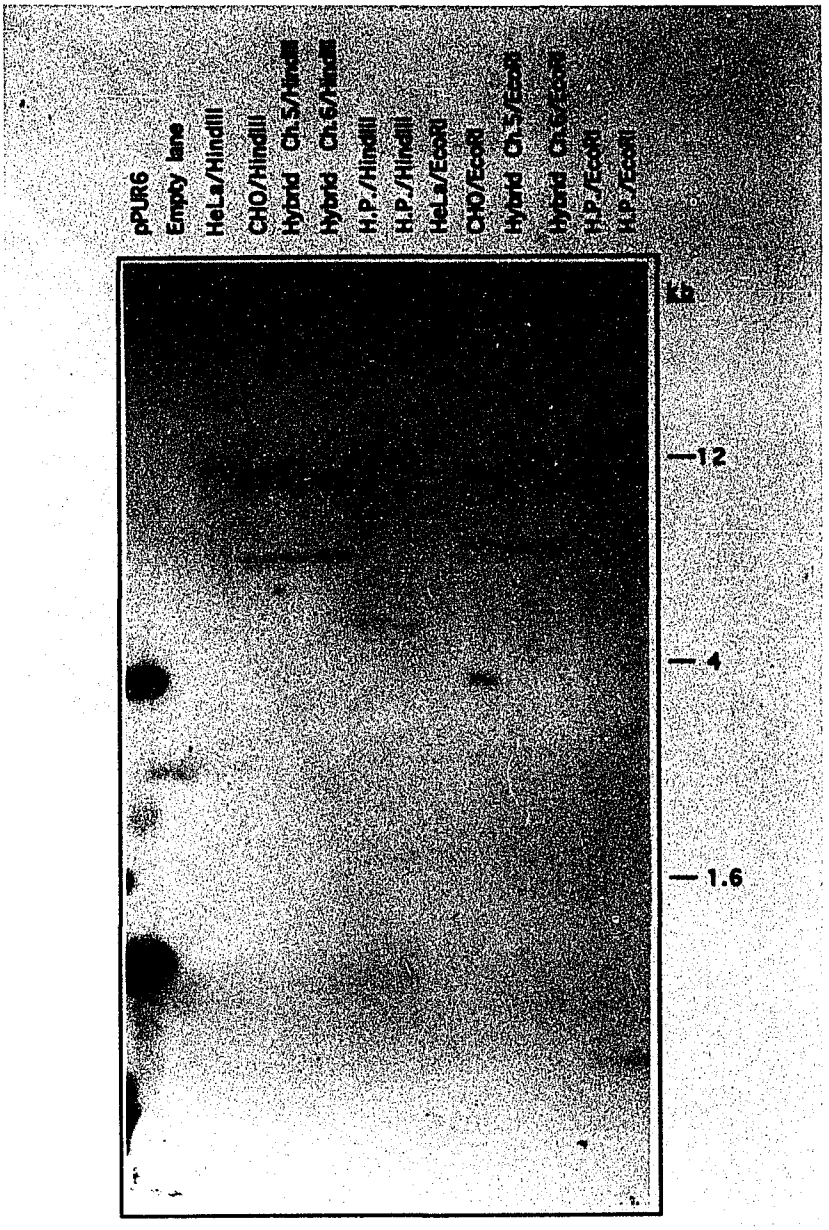

Fig. 17. Southern analysis of the *puro* gene with genomic DNA from human/hamster somatic hybrid cells. Five µg of genomic DNA from HeLa cells, or CHO cells, or human placenta (H.P.), or haman/hamster somatic hybrid cells containing single human chromosome 5 or chromosome 6 (hybrid Ch.5 or hybrid Ch.6) was

digested with individual restriction endonuclease (either *Hind*III, or *EcoR*I) as indicated at the top of the panel after the slash. The restriction fragments were separated on 0.8% agarose gel and transferred to GeneScreen Plus Membrane. Membranes were probed with labeled 777-bp Purα cDNA fragment as described in Materials and Methods. Abbreviation of pPUR6 indicates recombinant pUC19 DNA (10 pg) with a Purα cDNA insert, which was either digested with *Bam*HI to be linearized (4 kb) or digested with *EcoR*I to release the insert (1.1 kb). Numbers on the right show DNA molecular size markers in kilobases measured with the aligned 1 kb DNA ladder and λ DNA/*Hind*III fragments.

gene is located on chromosome 5 only. After integrating this result with the result of fluorescence *in situ* hybridization of the genomic *purα* gene on human chromosomes as described previously, we arrive at the conclusion that the human *purα* gene is located on 5q31.

The 4.5 kb *Hind*III fragment, which is from the *purβ* gene, weakly hybridized to the probe in the human control DNA samples, but was absent in either hybrid cell/chromosome 5 or 6. The 13 kb *EcoR*I fragment was present in the human control and chromosome 5 lanes, but absent in hamster control and hybrid cell/chromosome 6 lanes. Two additional *EcoR*I fragments from the human genome, 3 kb and 0.6 kb as shown in Fig. 8, gave relatively weak signals in Fig. 17. Genomic DNA from CHO cells gave one additional band in either the *Hind*III (the 11 kb band) or *EcoR*I digest (the 3.5 kb band) as compared with the correspondent lanes of the hamster hybrid cells (Fig. 17). There is no information available at this time to explain this phenomenon, although selective gene duplication in the CHO cells might be possible.

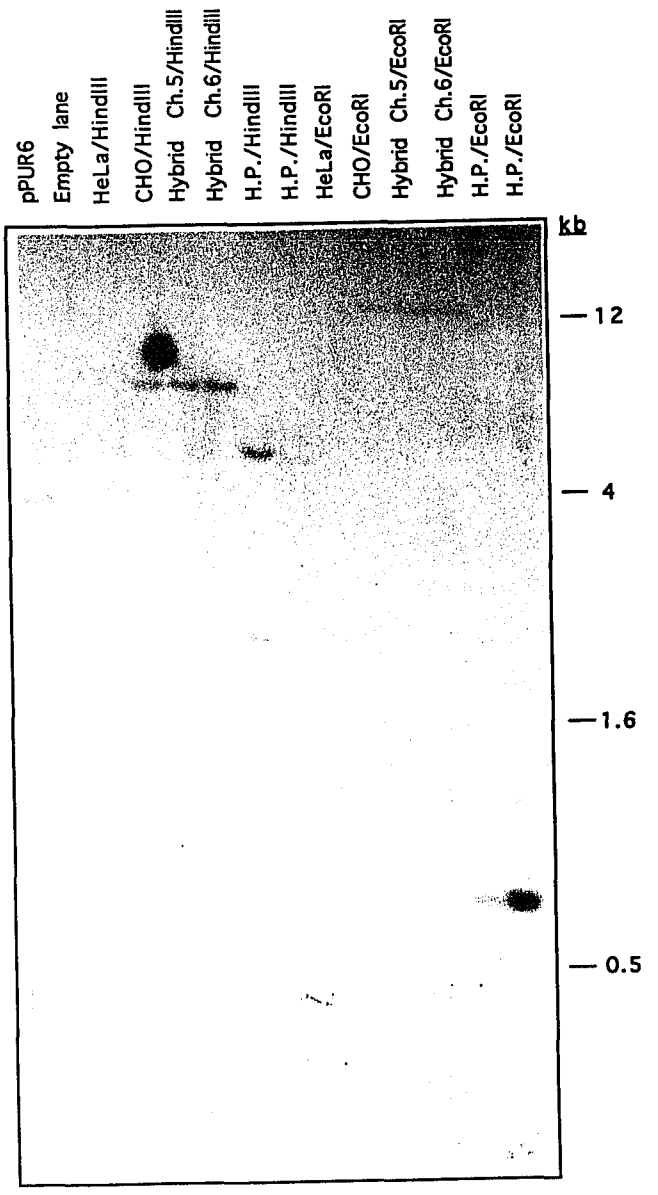

Fig. 18. Southern analysis of the *purβ* gene with genomic DNA from human/hamster somatic hybrid cells. Five μg of genomic DNA from HeLa cells, or CHO cells, or human placenta (H.P.), or haman/hamster somatic hybrid cells containing a single human

chromosome 5 or 6 (hybrid Ch.5 or hybrid Ch.6) was digested with individual restriction endonuclease (either *Hind*III, or *EcoR*I) as indicated at the top of the panel after the slash. The restriction fragments were separated on 0.8% agarose gel and transferred to GeneScreen Plus Membrane. Membranes were probed with labeled 406-bp Purβ cDNA fragment as described in Materials and Methods. Abbreviation of pPUR6 indicated recombinant pUC19 DNA (10 pg) with a Purα cDNA insert, which was either digested with *Bam*HI to be linearized (4 kb) or digested with *EcoR*I to release the insert (1.1 kb). Numbers on the right show DNA molecular size markers in kilobases measured with the aligned 1 kb DNA ladder and λ DNA/*Hind*III fragments.

G. The human purβ gene is located neither on chromosome 5 nor on chromosome 6

Genomic Southern blots with human/Chinese hamster somatic hybrid cells, which contains a single human chromosome either 5 or 6, were also performed to study the chromosomal localization of human *purβ* gene at the same conditions as that for *purα* gene in Fig. 17. The genomic DNAs was individually digested with either *Hind*III or *EcoR*I, and hybridized with the labeled human Purβ cDNA sequence, the 406 bp *EcoR*I fragment (Bergemann et al., 1992). The 4.5 kb and 2 kb *Hind*III bands were present in the human DNA control lanes (the lanes for HeLa and human placental DNA digests), but absent in hamster DNA control lane (CHO lane) and the lanes for hybrid cell/chromosome 5 or 6 (Fig. 18). The 0.6 kb *EcoR*I fragment was present in the human control DNA samples only, but absent in the hamster control lane and hybrid cell/chromosome 5 or 6. These results indicate that the

human *purβ* gene is located neither at chromosome 5 nor at chromosome 6.

4. Studies on the human Purα protein

Human Purα protein, expressed as a glutathione-S-transferase-Purα (GST-Purα) fusion protein, specifically binds to the hypophosphorylated form of the retinoblastoma protein, Rb, with an affinity as high as that of SV40 large T-antigen (Johnson et al., 1994). I sought to determine whether or not Rb binding to Purα would supershift the Purα-DNA band seen under standard conditions for gel shift assays with labeled oligonucleotide (MFO677) corresponding to the Purα recognition element. For this I employed purified GST-Purα, isolated by glutathione affinity chromatography as previously described (Johnson et al., 1994), and purified $p56^{RB}$, isolated as previously described (Hensey et al., 1994). The $p56^{RB}$ protein includes most of the C-terminus of $p110^{RB}$, including the portion of the protein responsible for binding to many other proteins, including SV40 large T-antigen. Fig. 19A shows results of a typical gel shift study using these two purified proteins. Under standard conditions, $p56^{RB}$ binds insignificantly to the Purα DNA recognition element, as shown in the lane labeled Rb. On the other hand, GST-Purα yields several bands corresponding to multimeric forms of the Purα protein complexed with DNA (lane Pur). when GST-Purα is cleaved with thrombin to separate Purα and GST moieties, GST does not bind

the labeled DNA at all (data not shown). A plot of the migration distance of ladder bands a, b and c, fits best to the assumption that they are multiples of 1, 2 and 3 GST-Purα molecules, respectively. When GST-Purα (53 kDa) is cleaved to yield Purα (34 kDa), the resulting gel band migrates approximately twice as fast as that for GST-Purα (data not shown). Therefore, we believe that band a, seen complexed to DNA in the Pur lane, represents monomeric GST-Purα. When GST-Purα and $p56^{RB}$ are added together to the labeled oligonucleotide, either with or without prior preincubation of the proteins, several alterations in the band-shift pattern are observed. Primarily, the intensity of all ladder bands of GST-Purα is enhanced, including the aggregate migrating near the wells. Band c appears much more intense at approximately the position of trimeric Purα. In addition, the fastest-migrating bands, a and b, are slightly shifted to a slower-migrating position. This is best visualized for band a in Fig. 19B. Bands in Fig 19B, corresponding to similar lanes in Fig. 19A, are autoradiographically exposed for much less time than those at Fig. 19A, and the shift in band a is more distinct. Such a slight shift in the bands for monomeric and dimeric Purα would not be expected if $p56^{RB}$ is present in a ternary complex with GST-Purα and DNA. Further evidence that $p56^{RB}$ is not in this complex is provided by observations that several anti-Rb antibodies tested did not produce supershift for any of the shifted Purα-DNA bands (data not shown). In fact, although

Purα binds both DNA and Rb, we have obtained no evidence that Purα binds the two simultaneously.

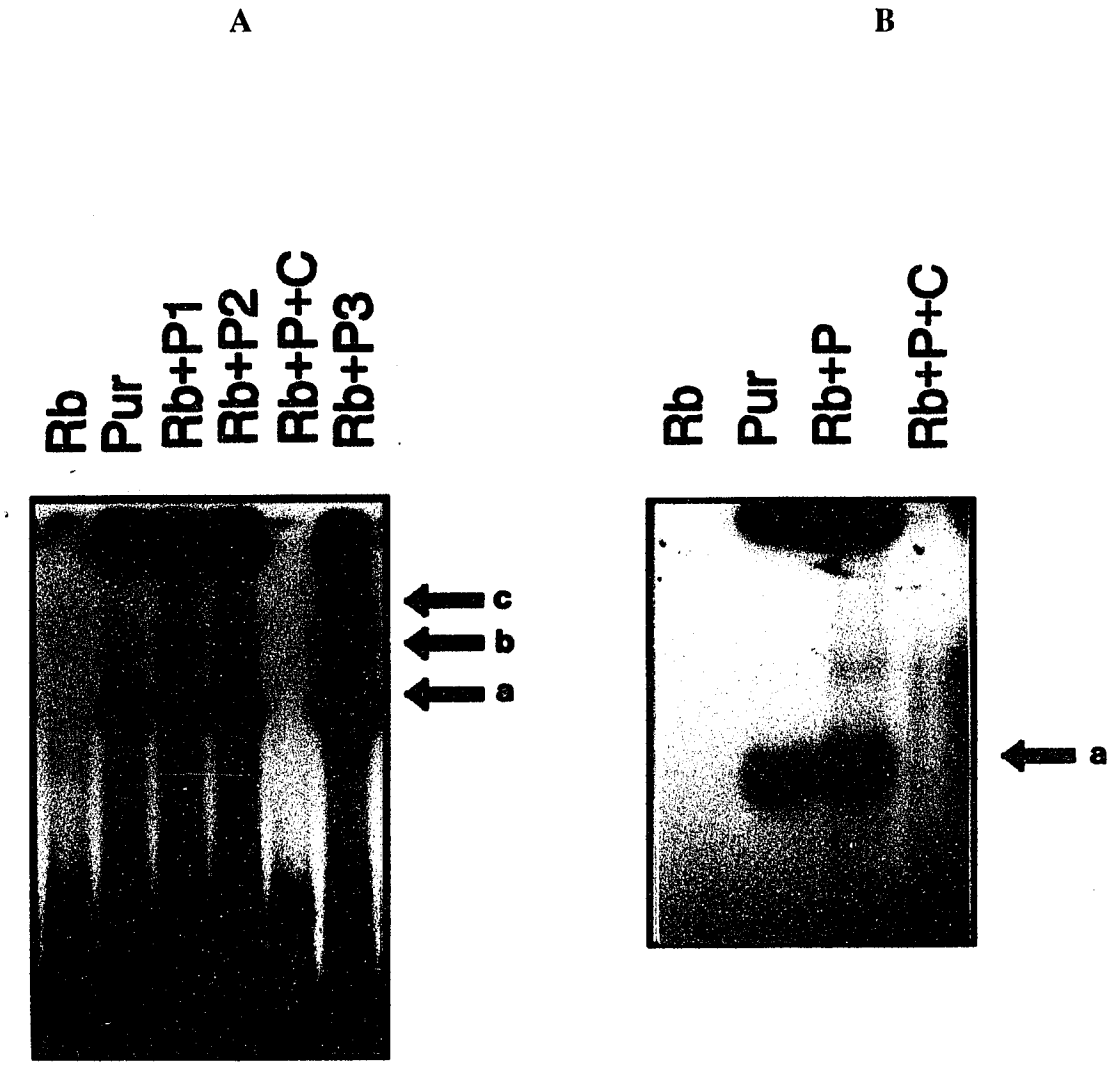

Fig. 19. Gel shift assay with GST-Purα and $p56^{RB}$. Abbreviations on the top panel are following. Rb: $P56^{RB}$, the purified 56 KDa form of

Rb, 50 ng; Pur: purified GST-Purα, 90 ng; Rb+P1, or Rb+P: GST-Purα (90 ng) + $P56^{RB}$ (50 ng) mixed directly before electrophoresis without a 15 minutes pre-incubation at 37°C; Rb+P2: GST-Purα (90 ng) + $P56^{RB}$ (50 ng) mixed and given the 15 minutes pre-incubation at 37°C prior to addition of labeled oligonucleotide and gel electrophoresis; Rb+P+C: GST-Purα (90 ng) + $P56^{RB}$ (50 ng) mixed and given the 15 minutes pre-incubation in the presence of 100-fold excess of unlabeled competitior MFO677 (150 ng); Rb+P3: GST-Purα (20 ng) + $P56^{RB}$ (50 ng) mixed and given the 15 minutes pre-incubation at 37°C prior to addition of labeled oligonucleotide and gel electrophoresis. Letters a, b and c at the right indicate the shifted bands.

CHAPTER 5

DISCUSSION

We have identified the cDNA sequence of Purα, a sequence-specific DNA-binding protein. Clues to the functions of Purα may be achieved by analysis of its predicted amino acid sequence, which may serve as primary guidelines for the functional studies of Purα. Knowledge of the Purα sequence will also facilitate the studies of the functional domains of Purα by genetic approaches. The ability of Purα to unwind duplex DNA in a sequence-specific manner is one such function that can be studied *in vitro* using mutants generated on the basis of sequence information.

The 3' and 5' ends of the human cDNA sequence presented in Fig. 2 agree very well with major bands observed upon 3' and 5' RACE of Purα. 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. Since there is an extraordinary degree of conservation between human Purα (hPurα) and mouse Purα (mPurα), and the mouse 5' untranslated sequence contains in-frame stop codons (Ma et al., 1994a; Fig. 6), it suggests that the predicted first methionine of hPurα is truly the first codon. 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 extracts (Bergemann and Johnson, 1992).

Considerable evidence now indicates that DNA replication initiates within discrete zones in mammalian genomes.

Although size estimates for such initiation zones vary from approximately 0.5 to 30 kb for various loci, there is strong implication that sequences within these zones serve some regulatory functions in initiation of DNA replication. There is no direct indication as yet that *cis*-acting sequences act as the binding sites for factors that control initiation of DNA replication, 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 (Eckdahl and Anderson, 1990). The binding of Purα to regions of stable bending DNA in five mammalian initiation zones makes this protein a strong candidate for such a *trans*-acting factor.

The predicted structure 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 (Haynes et al., 1987; Haynes et al., 1990; Steinert et al., 1991; Ge et al., 1991;). 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 (Roger et al., 1986). 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 (Ge et al., 1991).

Glutamine-rich regions and amphipathic helices such as those in the carboxyl terminus of Purα are common features of the transactivation domains of many transcription factors (Hope and Struhl, 1986; Ptashne, 1988; Courey et al., 1989). 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 transcription control elements (Hay et al., 1985). Binding sites for Purα appear within the promoter regions of several mammalian genes (Bergemann and Johnson, 1992). But in the hamster *dhfr* initiation zone the Purα elements are located approximately 17 kb 3' to the nearest known transcription unit. 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 transcription regulation and also interacts specifically with the adenoviral DNA polymerase during preinitiation of adenovirus DNA replication (Chen et al., 1990). The yeast *ARS*-binding factor I (ABF1) is involved in transcription activation, transcription silencing, and *ARS* activation (Rhode et al., 1992). 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 (Marahrens and Stillman, 1992).

The MCM proteins of yeast belong to a family of nucleic acid-dependent ATPases involved in the initiation of DNA replication in both prokaryotes and eukaryotes (Koonin, 1993). MCM2 is reported to be important for *ARS*-dependent initiation of DNA replication in yeast (Yang et al., 1991). Purα, which binds an element present at several recently-mapped mammalian origins of replication, has regions of homology to the ATP-binding motif of MCM proteins and of other proteins involved in initiation of replication, including the DnaA protein of *E. coli* (Koonin, 1993). BM28, a human MCM2 homolog, is reportedly required for entry of cell into S-phase (Todorov et al., 1994). These cellular proteins possess homologies to SV40 large T-antigen, a well-characterized replication initiation protein and transcription factor. The homology of Purα to T-antigen spans a region necessary for binding of T-antigen to the retinoblastoma protein, Rb (Ma et al., 1994a).

Rb is a nuclear phosphoprotein of 110 kDa which is constitutively expressed in most cultured cells (Lee et al., 1987) and normal tissues (Bernards et al., 1989). The phosphorylation of Rb protein oscillates during the cell division cycle. The hypophosphorylated forms predominate in the G0 and G1 phases of cell cycle while the hyperphosphorylated forms are predominate in the S, G2 and M phases. The phosphorylation status of Rb can affect its biological functions, such as its ability to bind DNA. The hypophosphorylated form

is likely to be the biologically active form (for review see Goodrich and Lee, 1993). Purα specifically binds to the hypophosphorylated form of Rb with an affinity as high as that of SV40 large T-antigen. Mutational analysis has indicated that Purα binds Rb at the putative amphipathic α-helix. Purα-Rb complex can be immuno-precipitated from extracts of HepG2 cells (Johnson et al., 1994).

When Rb is hypophosphorylated, it forms filamentous aggregates which are tethered to the inner regions of the nuclear membrane. In this state Rb may associated with Purα in cells, and this would limit access of Purα to many, perhaps all, of its recognition elements in chromosomes in a manner yet to be detailed. It is shown in Fig. 16 that Purα also forms multimeric aggregates and that this formation is promoted by Rb. Thus Rb essentially promotes a cooperative mechanism for sequestering Purα. The Purα-Rb complex is disrupted by Rb phosphorylation (Johnson et al., 1994) which is the case in the late G1 phase of cell cycle. This may desequester Purα from Rb. Purα has a 10-fold preference for binding to the purine-rich single-stranded to the double-stranded version of its recognition element (Bergemann and Johnson, 1992), and could thus be expected to open, or to maintain open, the DNA duplex at specific sites.

The complex between Purα and its binding element is altered by Rb even after Purα and Rb are no longer associated (Fig. 16). If this represents an effect of Rb on the folding of Purα, it could suggest a chaperone function for the Rb protein.

The details and the longevity of this effect have yet to be determined.

The human genomic clone LGT1 hybridized to 2 loci on human chromosomes may suggest that the *pur*α gene should contain a sequence fragment homologous to another gene, which is not the *pur*β gene. The homologous sequence fragment should be located outside the exon that containing the 777 bp *Pst*I fragment. It could be either other exons or the introns of the gene.

The human *pur*α gene has been mapped to the long arm of chromosome 5 at band 5q31. The location 5q31 is one of the most common fragile sites of human chromosomes, as indicated by population cytogenetic study of aphidicolin-induced fragile sites on cultured human lymphocytes (Tedeschi et al., 1992). Interstitial deletions of chromosome 5q31 have been reported to have a strong correlation with non-lymphocytic leukemia (Thornton et al., 1991; Pedersen and Jensen, 1991; Beau et al., 1993; Kroef et al., 1993). Recent work in our laboratory indicates that the *pur*α gene is mutated in human malignant mesothelioma cells (Gan et al., unpublished data). It will be very interesting to know whether the *pur*α gene mutation is responsible for the malignancy in cases of the 5q31 deletion and whether the *pur*α gene mutation is relevant to the development of mesothelioma.

CHAPTER 6

APPENDIX: PUBLICATIONS

Bergemann, A. D., **Ma, Z.-W.**, and Johnson, E. M. (1992) Sequence of cDNA comprising the human *pur* gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. Mol. Cell. Biol. **12**, 5673-5682.

Johnson, E. M., Chen, P.-L., Krachmarov, C. P., Barr, S. M., **Ma, Z.-W.**, and Lee, W.-H. Association of human Purα with the retinoblastoma protein, Rb, regulates binding to the Purα single-stranded DNA recognition element. Manuscript submitted for publication.

Ma, Z.-W., Bergemann, A. D., and Johnson, E. M. (1994) Conservation in human and mouse of a motif common to several proteins involved in initiation of DNA replication. Gene in press.

Ma, Z.-W., Najfeld, V., and Johnson, E. M. (1994) Chromosomal localization of the human *purα* gene to 5q31. Manuscript in preparation.

CHAPTER 7

BIBLIOGRAPHY

Alberts, B. M. (1990) Recipes for replication. Nature **346,** 514-515.

Anachkova, B. and Hamlin, J. L. (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.

Anderson, J. N. (1986) Detection, sequence patterns and function of unusual DNA structures. Nucleic Acids Res. **14,** 8513-8533.

Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983) Cell **34,** 779-787.

Beau, M. M. L., Espinosa, R. III, Neuman, W. L., Stock, W., Roulston, D., Larson, R. A., Keinanen, M. and Westbrook, C. A. (1993) Proc. Natl. Acad. Sci. USA **90,** 5484-5488.

Bell, S. P. and Stillman, B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature **357,** 128-134.

Bell, S. P., Kobayashi, R., Stillman, B. (1993) Yeast origin recognition complex functions in transcription silencing and DNA replication. Science **262,** 1844-1849.

Bergemann, A. D. and Johnson, E. M. (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.

Bergemann, A. D., Ma, Z.-W., and Johnson, E. M. (1992) Sequence of cDNA comprising the human *pur* gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. Mol. Cell. Biol. **12,** 5673-5682.

Bernards, R., Schackleford, G. M., Gerber, M. R., Horowitz, J. M., Friend, S. H., Schartl, M., Bogenmann, E., Rapaport, J. M., McGee, T., Dryja, T. P. and Weinberg, R. A. (1989) Structure and expression of the murine retinoblastoma gene and characterization of its encoded protein. Proc. Natl. Acad. Sci. USA **86,** 6474-6478.

Bird, R. E., Louarn, J., Martuscelli, J., and Caro, L. (1972) Origin and sequence of chromosome replication in *Escherichia coli*. J. Mol. Biol. **70**, 549-566.

Blumenthal, A. B., Kriegstein, H. J., and Hogness, D. S. (1973) The units of DNA replication in *Drosophila melanogaster* chromosomes. Cold Spring Harbor Symp. Quant. Biol. **38**, 205-223.

Borowiec, J. A., Dean, F. B., Bullock, P. A., and Hurwitz, J. (1990) Binding and unwinding-how T antigen engages the SV40 origin of DNA replication. Cell **60**, 181-184.

Bramhill, D. and Kornberg, A. (1988a) Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell **52**, 743-755.

Bramhill, D. and Kornberg, A. (1988b) A model for initiation at origins of DNA replication. Cell **54**, 915-918.

Brand, A. H., Micklem, G., and Nasmyth, K. (1987) A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. Cell **51**, 709-719.

Bravo, R. and Macdonald-Bravo, H. (1987) Existence of two populations of cyclin/proliferating cell nuclear antigen during the cell cycle: Association with DNA replication sites. J. Cell Biol. **105**, 1549-1554.

Brewer, B. J. and Fangman, W. L. (1987) The localization of replication origins on *ARS* plasmids in *S. cerevisiae*. Cell **51**, 463-471.

Brewer, B. J. and Fangman, W. L. (1988) A replication fork barrier at the 3' end of yeast ribosomal RNA genes. Cell **55**, 637-643.

Broach, J. R., Li, Y.-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K. A., and Hicks, J. B. (1982) Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp. Quant. Biol. **47**, 1165-1173.

Brown, J. A., Holmes, S. G., and Smith, M. M. (1991) The chromatin structure of *Saccharomyces cerevisiae* autonomously replicating sequences changes during the cell division cycle. Mol. Cell. Biol. **11,** 5301-5311.

Buchman, A. R. and Kornberg, R. D. (1990) A yeast *ARS*-binding protein activates transcription synergistically in combination with other weak activating factors. Mol. Cell. Biol. **10,** 887-897.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991) Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA **88**, 4005-4009.

Burhans, W. C., Selegue, J. E., and Heintz, N. H. (1986a). Isolation of the origin of replication associated with the amplified Chinese hamster dihydrofolate reductase domain. Proc. Natl. Acad. Sci. USA **83**, 7790-7794.

Burhans, W. C., Selegue, J. E., and Heintz, N. H. (1986b). Replication intermediates formed during initiation of DNA synthesis in methotrexate-resistant CHOC 400 cells are enriched for sequences derived from a specific, amplified restriction fragment. Biochem. **25**, 441-449.

Burhans, W. C., Vassilev, L., Caddle, M. S., Heintz, N. H., and DePamphilis, M. L. (1990) Identification of an origin of bidirectional DNA replication in mammalian chromosomes. Cell **62,** 955-965.

Caddle, M. S., Lussier, R. H., and Heintz, N. H. (1990a) Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. J. Mol. Biol. **211,** 19-33.

Caddle, M. S., Dailey, L. and Heintz N. H. (1990b) RIP60, a mammalian origin-binding protein, enhances DNA bending near the dihydrofolate reductase origin of replication. Mol. Cell. Biol. **10,** 6236-6243.

Callan, H. G. (1973) DNA replication in the chromosomes of eukaryotes. Cold Spring Harbor Symp. Quant. Biol. **38,** 195-203.

Camphell, J. L. (1986) Eukaryotic DNA replication. Annu. Rev. Biochem. **55,** 733-771.

Carmichael, E. P., Roome, J. M. and Wahl, A. (1993) Binding of a sequence-specific single-stranded DNA-binding factor to the Simian virus 40 core origin inverted repeat domain is cell cycle regulated. Mol. Cell. Biol. **13,** 408-420.

Carroll, S. M., DeRose, M. L., Kolman, J. L., Nonet, G. H., Kelly, R. E. and Wahl, G. M. (1993) Localization of a bidirectional DNA replication origin in the native locus and in episomally amplified murine adenosine deaminase loci. Mol. Cell. Biol. **13,** 2971-2981.

Celniker, S. E., Sweder, K., Srienc, F., Bailey, J. E., and Campbell, J. L. (1984) Deletion mutations affecting autonomously replicating sequence *ARS1* of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4,** 2455-2466.

Chen, M., Mermod, N. and Horwitz, M. S. (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.

Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986) A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. **6,** 4723-4733.

Chou, P. Y. and Fasman, G. D. (1974) Prediction of protein conformation. Biochem. **13,** 222-245.

Cook, P. R. (1991) The nucleoskeleton and the topology of replication. Cell **66,** 627-635.

Courey, A. J., Holtzman, D. A., Jackson, S. P. and Tjian, R. (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59,** 827-836.

Cox, L. S. and Laskey, R. A. (1991) DNA replication occurs at discrete sites in pseudonuclei assembled from purified DNA *in vitro*. Cell **66,** 271-275.

Dailey, L., Caddle, M. S., Heintz, N., and Heintz, N. H. (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.

Darnell, J. E., Jr. (1982) Variety in the level of gene control in eukaryotic cells. Nature **297,** 365-371.

DeCaprio, J. A., Ludlow, J. W., Figge, J., Shew, J.-Y., Huang, C.-M., Lee, W.-H., Marsillio, E., Paucha, E. and Livingston, D. M. (1988) SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. Cell **54,** 275-283.

De Cicco, D. V. and Spradling, A. C. (1984) Localization of a *cis*-acting element responsible for the developmentally regulated amplification of *Drosophila* chorion genes. Cell **38,** 45-54.

Delidakis, C. and Kafatos, F. (1989) Amplification enhancers and replication origins in the autosomal chorion gene cluster of *Drosophila*. EMBO J. **8,** 891-901.

DePamphilis, M. L. (1988) Transcriptional elements as components of eukaryotic origins of DNA replication. Cell **52,** 635-638.

DePamphilis, M. L. (1993) Eukaryotic DNA replication: Anatomy of an origin. Annu. Rev. Biochem. **62,** 29-63.

DePinho, R. A., Legony, E., Feldman, L. B., Kohl, N. E., Yancopoulos, G. D., and Alt, F. W. (1986) Structure and expression of the murine *N-myc* gene. Proc. Natl. Acad. Sci. USA **83,** 1827-1831.

Diffley, J. F. X. and Stillman, B. (1988) Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. Proc. Natl. Acad. Sci. USA **85,** 2120-2124.

Diffley, J. F. X. and Stillman, B. (1990) The initiation of chromosomal DNA replication in eukaryotes. Trends Genet. **6,** 427-432

Dijkwel, P. A., Vaughn, J. P., and Hamlin, J. L. (1991) Mapping of replication initiation sites in mammalian genomes by two-dimensional gel analysis: Stabilization and enrichment of replication intermediates by isolation on the nuclear matrix. Mol. Cell. Biol. **11,** 3850-3859.

Dobbs, D. L., Shaiu, W.-L. and Benbow, R. M. (1994) Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. in press.

Dubey, D. D. and Raman, R. (1987) Factors influencing replicon organization in tissues having different S-phase durations in the mole rat, *Bandicota bengalensis*. Chromosoma **95,** 285-289.

Dubey, D. D., Davis, L. R., Greenfeder, S. A., Ong, L. Y., Zhu, J., Broach, J. R., Newlon, C. S., and Huberman, J. A. (1991) Evidence suggesting that the *ARS* elements associated with silencers of the yeast mating-type locus *HML* do not function as chromosomal DNA replication origins. Mol. Cell. Biol. **11,** 5346-5355.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M., and Levens, D. (1994) A sequence-specific, single-strand binding protein activates the far upstream element of *c-myc* and defines a new DNA-binding motif. Genes Dev. **8,** 465-480.

Eckdahl, T. T., and Anderson, J. N. (1987) Computer modelling of DNA structures involved in chromosome maintenance. Nucleic Acids Res. **15,** 8531-8545.

Eckdahl, T. T., and Anderson, J. N. (1990) Conserved DNA structures in origins of replication. Nucleic Acids Res. **18,** 1609-1612.

Edenberg, H. J. and Huberman, J. A. (1975) Eukaryotic chromosome replication. Ann. Rev. Genet. **9,** 245-284.

Eisenberg, S., Civalier, C., and Tye, B.-W. (1988) Specific interaction between a *Saccharomyces cerevisiae* protein and a DNA element associated with certain autonomously replicating sequences. Proc. Natl. Acad. Sci. USA **85,** 743-746.

Erdile, L. F., Heyer, W. D., Kolodner, R., and Kelly, T. J. (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.

Fangman, W. L. and Brewer, B. J. (1991) Activation of replication origins within yeast chromosomes. Annu. Rew. Cell Biol. **7,** 375-402.

Ferguson, B. M. and Fangman, W. L. (1992) A position effect on the time of replication origin activation in yeast. Cell **68,** 333-339.

Fort, P., Marty, L., Piechaczyk, M., Sabrouty, S. E., Dani, C., Jeanteur, P., and Blanchard, J. M. (1985) Various rat adult tissues express only one major mRNA species from the glyceraldehyde-3-phosphate-dehydrogenase multigenic family. Nucleic Acids Res. **13,** 1431-1442.

Foss, M., McNally, F. J., Laurenson, P. and Rine J. (1993) Origin recognition complex (ORC) in transcription silencing and DNA replication in *S. cerevisiae*.Science **262,** 1838-1844.

Francesconi, S. C. and Eisenberg, S. (1989) Purification and characterization of OBF1: a *Saccharomyces cerevisiae* protein that binds to autonomously replicating sequences. Mol. Cell. Biol. **9,** 2906-2913.

Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170,** 827-842.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85,** 8998-9002.

Gale, G. M., Tobey, R. A. and D'Anna, J. A. (1992) Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells. J. Mol. Biol. **244,** 343-358.

Ge, H., Zuo, P. and Manley, J. L. (1991) Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulatos. Cell **66,** 373-382.

Goldman, M. A., Hoimquist, G. P., Gray, M. C., Caston, L. A., and Nag, A. (1984) Replication timing of genes and middle repetitive sequences. Science **224,** 686-692.

Goodrich, D. W. and Lee, W.-H. (1993) Molecular characterization of the retinoblastoma susceptibility gene. Biochim. Biophy. Acta **1155,** 43-61.

Graesser, F., Scheidtmann, K. H., Tuazon, P. T., Traugh, J. A. and Walter, G. (1988) *In vitro* phosphorylation of SV40 large T antigen. Virology **165,** 13-22.

Hand, R. (1978) Eucaryotic DNA: Organization of the genome for replication. Cell **15,** 317-325.

Handeli, S., Klar, A. Meuth, M., and Cedar, H. (1989) Mapping replication units in animal cells. Cell **57,** 909-920.

Hay, N., Bishop, J. M. and Levens, D. (1985) Regulatory elements that modulate expression of human *c-myc*. Genes Dev. **1,** 659-671.

Haynes, S. R., Rebbert, M. L., Mozer, B. L., Forquignon, F and Dawid, I. B. (1987) Pen repeat sequences are GGN clusters and encode a glycine-rich domain in a *Drosophila* cDNA homologous to the rat helix desdabilizing protein. Proc. Natl. Acad. Sci. USA **84,** 1819-1823.

Haynes, S. R., Raychaudhuri, G. and Beyer, A. L. (1990) The *Drosophila* Hrb98DE locus encodes four protein isoforms homologous to the A1 protein of mammalian heterogeneous nuclear ribonucleoprotein complex. Mol. Cell. Biol. **10,** 316-323.

Heck, M. M. S. and Spradling, A. C. (1990) Multiple replication origins are used during *Drosophila* chorion gene amplification. J. Cell Biol. **110,** 903-914.

Heintz, N. H. and Hamlin, J. L. (1982) An amplified chromosomal sequence that includes the gene for dihydrofolate reductase

initiates replication within specific restriction fragments. Proc. Natl. Acad. Sci. USA **79,** 4083-4087.

Heintz, N. H., Milbrandt, J. D., Greisen, K. S., and Hamlin, J. L. (1983) Cloning of the initiation region of a mammalian chromosomal replicon. Nature **302,** 439-441.

Heintz, N. H. (1992) Transcription factors and the control of DNA replication. Curr. Opin. Cell Biol. **4,** 459-467.

Heintz, N. H., Dailey, L., Held, P. and Heintz, N. (1992) Eukaryotic replication origins as promoters of bidirectional DNA synthesis. Trends in Genet. **8,** 376-381.

Hensey, C. E., Hong, F., Durfee, T., Qian, Y.-W., Lee, E. Y.-H. P., and Lee, W.-H. (1994) Identification of discrete structural domains in the retinoblstoma protein. J. Biol. Chem. **269,** 1380-1387.

Herrick, J. and Bensimon, D. (1991) Gene regulation under growth conditions. Amodel for the regulation of replication in *Escherichia coli.* J. Theori. Biol. **151,** 359-366.

Hofmann, J. F.-X. and Gasser, S. M. (1991) Identification and purification of a protein that binds the yeast *ARS* consensus sequence. Cell **64,** 951-960.

Hope, I. A. and Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell **46,** 885-894.

Houten, J. V. V. and Newlon, C. S. (1990) Mutational analysis of the consensus sequence of a replication origin from yeast chromosome III. Mol. Cell. Biol. **10,** 3917-3925.

Houck, C. M., Rinehart, F. P., and Schmid, C. W. (1979) A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. **132,** 289-306.

Huberman, J. A. and Riggs, A. D. (1968) On the mechanism of DNA replication in mammalian chromosomes. J Mol. Biol. **32,** 327-341.

Huberman, J. A., Spotila, L. D., Nawotka, K. A., El- Assouil, S. M., and Davis, L. R. (1987) The *in vivo* replication origin of the yeast 2μm plasmid. Cell **51,** 473-481.

Huberman, J. A., Zhu, J., Davis, L. R., and Newlon, C. S. (1988) Close association of a DNA replication origin and an *ARS* element on chromosome III of the yeast, *Saccharomyces cerevisiae*. Nucleic Acids Res. **16,** 6373-6384.

Hurwitz, J., Dean, F. B., Kwong, A. D. and Lee, S.-H. (1990) The in vitro replication of DNA containing the SV40 origin. J. Biol. Chem. **265,** 18043-18046.

Hutchison, C. and Kill, I. (1989) Changes in the nuclear distribution of DNA polymerase alpha and PCNA/cyclin during the progress of the cell cycle, in a cell-free extract of *Xenopus* eggs. J. Cell Sci. **93,** 605-613.

Iguchi-Ariga, S. M. M., Okazaki, T., Itani, T., Ogata, M., Sato, Y., and Ariga, H. (1988) An initiation site of DNA replication with transcriptional enhancer activity present upstream of the *c-myc* gene. EMBO J. **7,** 3135-3142.

Jackson, D. A. (1990) The organization of replication centers in higher eukaryotes. BioEssays **12,** 87-89.

Johnson, E. M., Chen, P.-L., Krachmarov, C. P., Barr, S. M., Ma, Z.-W., and Lee, W.-H. Association of human Purα with the retinoblastoma protein, Rb, regulates binding to the Purα single-stranded DNA recognition element. Manuscript submitted for publication.

Kearsey, S. (1984) Structural requirements for the function of a yeast chromosomal replicator. Cell **37,** 299-307.

Kenelly, P. J. and Krebs, E. G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. J. Biol. chem. **266,** 15555-15558.

Kim, C., Snyder, R. O. and Wold, M. S. (1992) Binding properties of replication protein A from human and yeast cells. Mol. Cell. Biol. **12,** 3050-3059.

Kitsberg, D., Selig, S., Keshet, I. and Cedar, H. (1993) Replication structure of the human β-globin gene domain. Nature **366,** 588-590.

Koonin, E. V. (1993) A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acids Res. **11,** 2541-2547.

Kornberg, A. (1988) DNA replication. J. Biol. Chem. **263,** 1-4.

Kowalski, D. and Eddy, M. J. (1989) The DNA unwinding element: a novel, *cis*-acting component that facilitates opening of the *Escherichia coli* replication origin. EMBO J. **8,** 4335-4344.

Kriegstein, H. J. and Hogness, D. S. (1974) Mechanism of DNA replication in *Drosophila* chromosomes: structure of replication forks and evidence for bidirectionality. Proc. Natl. Acad. Sci. USA **71,** 135-139.

Kroef, M. J. P. L., Fibbe, W. E., Mout, R., Jansen, R. P. M., Haak, H. L., Wessels, J. W., Kamp, H. V., Willemze, R. and Landegent, J. E. (1993) Myeloid but not lymphoid cells carry the 5q deletion: polymerase chain reaction analysis of loss of heterozygosity using mini-repeat sequences on highly purified cell fractions. Blood **81,** 1849-1854.

Krysan, P. J., Haase, S. B., and Calos, M. P. (1989) Isolation of human sequence that replicate autonomously in human cells. Mol. Cell. Biol. **9,** 1026-1033.

Kuno, K., Murakami, S., and Kuno, S. (1990) Single-strand-binding factor(s) which interact with *ARS1* of *Saccharomyces cerevisiae*. Gene **95,** 73-77.

Labarca, C. and Paigen, K. (1980) A simple, rapid, and sensitive DNA assay procedure. Anal. Biochem. **102,** 344-352.

Laskey, R. A., Fairman, M. P., and Blow, J. J. (1989) S phase of the cell cycle. Science **246,** 609-614.

Lee, W.-H., Shew, J.-Y., Hong, F. D., Sery, T. W., Donoso, L. A., Young, L.-J., Bookstein, R. and Lee, E. Y.-H. (1987) The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. Nature **329,** 642-645.

Leu, T.-H. and Hamlin, J. L. (1989) High-resolution mapping of replication fork movement through the amplified dihydrofolate reductase domain in CHO cells by in-gel renaturation analysis Mol. Cell. Biol. **9,** 523- 531.

Leu, T.-H., Anachkova, B., and Hamlin, J. L. (1990) Repetitive sequence elements in an initiation locus of the amplified dihydrofolate reductase domain in CHO cells. Genomics **7,** 428-433.

Linskens, M. H. K. and Huberman, J. A. (1988) Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8,** 4927-4935.

Linskens, M. H. K. and Huberman, J. A. (1990) Ambiguities in results obtained with 2D gel replicon mapping techniques. Nucleic Acids Res. **18,** 647-652.

Ma, Z.-W., Bergemann, A. D., and Johnson, E. M. (1994a) Conservation in human and mouse of a motif common to several proteins involved in initiation of DNA replication. Gene in press.

Ma, Z.-W., Najfeld, V., and Johnson, E. M. (1994b) Chromosomal localization of the human purα gene to 5q31. Manuscript in preparation.

Maine, G. T., Sinha, P., and Tye, B-K. (1984) Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. Genetics **106,** 365-385.

Marahrens, Y. and Stillman, B. (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science **255,** 817-823.

Marraccino, R. L., Fotedar, R., D'Urso, G., and Roberts, J. M. (1990) Control of DNA replication. Curr. Opin. Cell Biol. **2,** 262-268.

Mastrangelo, I. A., Hough, P. V. C., Wall, J. S., Dodson, M., Dean, F. B., and Hurwitz, J. (1989) ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. Nature **341**, 658-662.

Mastrangelo, I. A., Held, P. G., Dailey, L., Wall, J. S., Hough, P. V. C., Heintz, N. and Heintz, N. H. (1993) RIP60 dimers and multiples of dimers assemble link structures at an origin of bidirectional replication in the dihydrofolate reductase amplicon of Chinese hamster ovary cells. J. Mol. Biol. **232,** 766-778.

McWhinney, C. and Leffay, M. (1990) Autonomous replication of a DNA fragment containing the chromosomal replication origin of the human *c-myc* gene. Nucleic Acids Res. **18,** 1233-1242.

Milbrandt, J. D., Heintz, N. H., White, W. C., Rothman, S. M., and Hamlin, J. L. 1981. Methotrexate-resistant Chinese hamster ovavy cells have amplified a 135-kilobase-pair region that includes the dihydrofolate reductase gene. Proc. Natl. Acad. Sci. USA **78,** 6043-6047.

Mills, A. D., Blow, J. J., White, J. G., Amos, W. B., Wilcock, D., and Laskey, R. A. (1989) Replication occurs at discrete foci spaced throughout nuclei replicating *in vitro*. J. Cell Sci. **94,** 471- 477.

Najfeld, V., Ballard, S. G., Menninger, J., Ward, D. C., Bouhassira, E. E., Schwartz, R. S., Nagel, R. L. and Rybicki, A. C. (1992) The gene for human erythrocyte protein 4.2 maps to chromosome 15q15. Am. J. Hum. Genet. **50,** 71-75.

Nakamura, H., Morita, T., Masaki, S., and Yoshida, S. (1984) Intracellular localization and metabolism of DNA polymerase α in human cells visualized with monoclonal antibody. Exp. Cell Res. **151,** 123-133.

Nakamura, H., Morita, T., and Sato, C. (1986) Structural organization of replicon domains during DNA synthetic phase in the mammalian nucleus. Exp. Cell Res. **165,** 291-197.

Nakayasu, H. and Berezney, R. (1989) Mapping replicational sites in the eucaryotic cell nucleus. J. Cell Biol. **108,** 1-11.

Nelson-Rees, W. A., Hunter, L., Darlington, G. J. and O'Brien, S. J. (1980) Characteristics of HeLa strains: permanent vs. variable features. Cytogenet. Cell Genet. **27,** 216-231.

Orr-Weaver, T. L. and Spradling, A. C. (1986) *Drosophila* chorion gene amplification requires an upstream region regulating *s*18 transcription. Mol. Cell. Biol. **6,** 4624-4633.

Pedersen, B. and Jensen, I. M. (1991) Clinical and prognostic implications of chromosome 5q deletions: 96 high resolution studied patients. Leukemia **5,** 566-573.

Prescott, D. M. and Kuempel, P. L. (1972) Bidirectional replication of the chromosome in *Escherichia coli.* Proc. Natl. Acad. Sci. USA **69,** 2842-2845.

Ptashne, M. (1988) How eukaryotic transcriptional activators work. Nature **335,** 683-689.

Rao, V. N., Papas, T. S. and Reddy, E. S. P. (1987) *erg*, a human *ets*-related gene on chromosome 21: Alternative splicing, polyadenylation and translation. Science **237,** 635-639.

Rhode, P. R., Elsasser, S. and Campbell, J. L. (1992) Role of multifunctional autonomously replicating sequence binding factor I in the initiation of DNA replication and transcription control in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **12,** 1064-1077.

Rivier, D. H. and Rine, J. (1992) An origin of DNA replication and a transcription silencer require a common element. Science **256,** 659-663.

Rost, B. and Sander, C. (1993) Improved prediction of protein secondary structure by use of sequence profiles and neural networks. Proc. Natl. Acad. Sci. USA **90,** 7558-7562.

Rogers, S., Wells, R. and Rechsteiner, M. (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science **234,** 364-368.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989a) Electrophoresis of RNA through gels containing formaldehyde. In Molecular Cloning, Volume 1, 2nd edition, Nolan, C., ed. (Cold Spring Harbor Laboratory Press), P 7.43-7.45.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989b) Small-scale preparation of plasmid DNA. In Molecular Cloning, Volume 1, 2nd edition, Nolan, C., ed. (Cold Spring Harbor Laboratory Press), P 1.29-1.30.

Sanger , F., Nicklen, S. and Coulson, A. R. (1977) DNA sequenceing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463-5467.

Schmidt, A. M. A., Herterich, S. U., and Krauss, G. (1991) A single-stranded DNA binding protein from *S. cerevisiae* specifically recognizes the T-rich strand of the core sequence of ARS elements and discriminates against mutant sequences. EMBO J. **10,** 981-985.

Schneikert, J., Lutz, Y. and Wasylyk, B. (1992) Two independent activation domains in c-Ets-1 and c-Ets-2 located in non-conserved sequences of the *ets* gene family. Oncogene **7,** 249-156.

Siddique, H. R., Rao, V. N., Lee, L. and Reddy, S. P. (1993) Characterization of the DNA binding and translational activation domains of the Erg protein. Oncogene **8,** 1751-1755.

Siebenlist, U. L., Hennighausen, L., Battey, J., and Leder, P. (1984) Chromatin structure and protein binding in the putative regulatory region of the *c-myc* gene in Burkitt lymphoma. Cell **37,** 381-391.

Simpson, R. T. (1990) Nucleosome positioning can affect the function of a *cis*-acting DNA element *in vivo*. Nature **343,** 387-389.

Snyder, M., Sapolsky, R. J., and Davis, R. W. (1988) Transcription interferes with elements important for chromosome maitenance in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8**, 2184-2194.

Spradling, A. C. (1981a) The organization and amplification of two chromosomal domains containing *Drosophila* chorion genes. Cell **27**, 193-201.

Spradling, A. C. (1981b) A chromosome inversion alters the pattern of specific DNA replication in *Drosophila* follicle cells. Cell **27**, 203-209.

Stanton, L. W., Schwab, M., and Bishop, J. M. (1986) Nucleotide sequence of the human *N-myc* gene. Proc. Natl. Acad. Sci. USA **83**, 1772-1776.

Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S. Q., Haynes, S. R. and Stevens, S. C. (1991) Glycine loops in proteins: their occurrence in certain intermediate filament chain, loricins and single-stranded RNA binding proteins. Int. J. Biol. Macromol. **13**, 130-139.

Stenzel, T. T., Pankaj P., and Bastia, D. (1987) The integration host factor of Escherichia coli binds to bent DNA at the origin of replication of the plasmid pSC101. Cell **49**, 709-717.

Stillman, B. (1989) Initiation of eukaryotic DNA replication in vitro. Annu. Rev. Cell Biol. **5**, 197-245.

Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979) Isolation and characterisation of a yeast chromosomal replicator. Nature **282**, 39-43.

Taylor, J. H. (1984) Origins of replication and gene regulation. Mol. Cell. Biochem. **61**, 99-109.

Tedeschi, B., Bernole, P., Sanna, M. L. and Nicoletti, B. (1992) Population cytogenetics of aphidicolin-induced fragile sites. Hum. Genet. **89**, 543-547.

Thornton, D. E., Theil, K., Payson, R., Balcerzak, S. P. and Chiu, I.-M. (1991) Characterization of the 5q- breakpoint in an acute

nonlymphocytic leukemia patient using pulsed-field gel electrophoresis. Am. J. Med. Genet. **41,** 557-565.

Todorov, I. T., Pepperkok, R. and Philopova, R. N. (1994) A human nuclear protein with sequence homology to a family of early S phase proteins is required for entry into S phase and for cell division. J. Cell Sci. in press

Umek, R. M. and Kowalski, D. (1988) The ease of DNA unwinding as a determinant of initiation at yeast replication origins. Cell **52,** 559-567.

Umek, R. M., Linskens, M. H. K., Kowalski, D., and Huberman, J. A. (1989) New beginnings in studies of eukaryotic DNA replication origins. Biochim. Biophys. Acta **1007,** 1-14.

Vassilev, L. and Russev, G. (1988) Purification of nascent DNA chains by immunoprecipitation with anti-BrdU antibodies. Nucleic Acids Res. **16,** 10397.

Vassilev, L. and Johnson, E. M. (1989) Mapping initiation sites of DNA replication *in vivo* using polymerase chain reaction amplification of nascent strand segments. Nucleic Acids Res. **17,** 7693-7705.

Vassilev, L. and Johnson, E. M. (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.

Vassilev, L., Burhans, W. C., and DePamphilis, M. L. (1990) Mapping an origin of DNA replication at a single-copy locus in exponentially proliferating mammalian cells. Mol. Cell. Biol. **10,** 4685-4689.

Vaughn, J. P., Dijkwel, P. A., and Hamlin, J. L. (1990) Replication initiation in a broad zone in the amplified CHO dihydrofolate reductase domain. Cell **61,** 1075-1087.

Virshup, D. M. (1990) DNA replication. Curr. Opin. Cell Biol. **2,** 453-460.

Virta-Pearlman, V. J., Gunaratne, P. H., and Chinault, A. C. (1993) Analysis of a replication initiation sequence from the

adenosine deaminase region of the mouse genome. Mol. Cell. Biol. **13,** 5913-5942.

Walker, F. E., Saraste, M., Runswick, R. J. and Gay, N. J. (1982) Distantly related sequences in the α- and β-subunits of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1,** 945-951.

Walker, S. S., Francesconi, S. C., and Eisenberg, S. (1990) DNA replication enhancer in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **87,** 4665-4669.

Wang, C.-Y., Petryniak, B., Thompson, C. B., Kaelin, W. G. and Leiden, J. M. (1993) Regulation of the Ets-related transcription factor Elf-1 by binding to the retinoblastoma protein. Science **260,** 1330-1335.

Wobbe, C. R., Weissbach, L., Borowiec, J. R., Dean, F. B., Murakami, Y., Bullock, P., and Hurwitz, J. (1987) Replication of simian virus 40 origin-containing DNA *in vitro* with purified proteins. Proc. Natl. Acad. Sci. USA **84,** 1834-1838.

Wold, M. D. and Kelly, T. (1988) Purification and characterization of replication protein A, a cellular protein required for *in vitro* replication of simian virus 40 DNA. Proc. Natl. Acad. Sci. USA **85,** 2523-2527.

Wu, H.-M., and Crothers, D. M. (1984) The locus of sequence-directed and protein-induced DNA binding. Nature **308,** 509-513.

Yang, H., Gibson, S. and Tye, B. K. (1991) MCM2 and MCM3, two proteins important for ARS activity, are related in structure and function. Genes Dev. **5,** 944-957.

Zahn, K., and Blattner, F. R. (1987) Direct evidence for DNA bending at the Lambda replication origin. Science **236,** 416-422.