MOLECULAR AND CELLULAR BIOLOGY, Jan. 1988, p. 321–329
0270-7306/88/010321-09$02.00/0

Vol. 8, No. 1

The C-Terminal Domain of the Largest Subunit of RNA Polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and Mammals: a Conserved Structure with an Essential Function

LORI A. ALLISON, JERRY K.-C. WONG, V. DANIAL FITZPATRICK, MATTHEW MOYLE, AND C. JAMES INGLES*

Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada M5G 1L6

Received 20 July 1987/Accepted 21 September 1987

Using DNA encoding the largest subunit of *Drosophila melanogaster* RNA polymerase II, we isolated the homologous hamster *RPO21* gene. Nucleotide sequencing of both the hamster and *D. melanogaster RPO21* DNAs confirmed that the *RPO21* polypeptides of these two species, like the *Saccharomyces cerevisiae RPO21* polypeptide, contain both an N-terminal region homologous to the *Escherichia coli* RNA polymerase subunit β′ and a unique polymerase II-specific C-terminal domain. This C-terminal domain, encoded by separate exons in the *D. melanogaster* and hamster genes, consists of a tandemly repeated heptapeptide sequence. By constructing a series of deletions in DNA encoding the 26 heptapeptide repeats normally present in the *S. cerevisiae RPO21* polypeptide, we have established that a minimum of between 9 and 11 repeats is necessary for *RPO21* function in yeast cells. Replacement of the yeast *RPO21* heptapeptide repeats by the longer hamster repetitive domain resulted in viable yeast cells with no detectable mutant phenotype, while a similar replacement of the yeast repeats by the more divergent *D. melanogaster* repeats was a recessive lethal mutation. We suggest that this novel repetitive domain is essential for proper initiation of transcription by RNA polymerase II and that it may mediate the functions of TATA boxes, upstream activating sequences, and enhancers.

Biochemical and genetic approaches are being used to study the molecular mechanisms which regulate mRNA transcription in eucaryotes. Through such approaches, regulatory *cis*-acting DNA elements have been identified and protein factors that recognize these specific sequences have been described. These DNA-binding transcription factors must directly or indirectly affect the activity of RNA polymerase II. To address this aspect of transcriptional regulation, a more thorough understanding of this enzyme itself may be required. Toward this end, we and others have used mutations that increase the α-amanitin resistance of RNA polymerase II to select, identify, or characterize other RNA polymerase II mutations (11, 15, 20). An α-amanitin resistance mutation in RNA polymerase II has also been used for the chromosomal mapping (11) and molecular cloning (25) of DNA encoding the largest subunit (10) of *Drosophila melanogaster* RNA polymerase II.

Previous studies in this laboratory showed that the RNA polymerase II DNA from the *D. melanogaster RPO21* (also called *RpII215* in reference 2) locus hybridized to mammalian DNA. Using DNA-mediated gene transfer, we identified cross-hybridizing restriction fragments in DNA from Chinese hamster ovary (CHO), Syrian hamster, and human cells that encode at least part of the mammalian RNA polymerase II polypeptide (16). Subsequently this *D. melanogaster* DNA was used to isolate the analogous polymerase II genes from *Saccharomyces cerevisiae* (17), human (4), and mouse (5) cells. Nucleotide-sequencing studies of the yeast genes *RPO21* and *RPO31* (1a), which encode the largest subunit of RNA polymerase II (1a, 17) and RNA polymerase III (21), respectively, and a similar analysis of a portion of the *D. melanogaster RPO21* DNA (2) established that the eucaryotic RNA polymerase polypeptides are remarkably similar in their primary structure to the analogous subunit of the *Escherichia coli* RNA polymerase, β′. In addition, the largest subunit of yeast (1a) and mouse (5) RNA polymerase II was found to contain a C-terminal extension. This domain consists of an evolutionarily conserved, tandemly repeated heptapeptide sequence with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser.

In the present study we report the isolation of DNA encoding the Chinese hamster RNA polymerase II *RPO21* polypeptide and the nucleotide sequences of the 3′ portion of both this hamster gene and the *D. melanogaster* gene. Like the *RPO21* polypeptide of *S. cerevisiae* and mouse cells, the hamster and *D. melanogaster* polypeptides have the unique eucaryotic-specific C-terminal domain. By making a series of deletions in the yeast *RPO21* heptapeptide domain and by expressing chimeric yeast-hamster and yeast-*D. melanogaster RPO21* genes in yeast cells, we have demonstrated that a minimum number of the heptapeptide repeats is essential in vivo and that the function of these heptapeptide repeats is in part species specific. We suggest that this domain of RNA polymerase II may provide a special set of sites for protein-protein contacts regulating mRNA synthesis.

MATERIALS AND METHODS

Isolation of Chinese hamster *RPO21* DNA. Three libraries (each containing 5×10^5 independent isolates) of hamster cell genomic DNA from the α-amanitin-resistant transformant Tr-pGA-Ama1-5-6 (16) were constructed in λEMBL vectors (8). From the first library containing partial *MboI*-cut genomic DNA, we isolated the single bacteriophage λAma A-1 by using the *D. melanogaster RPO21* DNAs p4.2 and p4.1 (25) as hybridization probes at low stringency (16). Portions of λAma A-1 were used in turn to isolate two additional phages, λAma D-1 and λAma E10, from libraries

* Corresponding author.

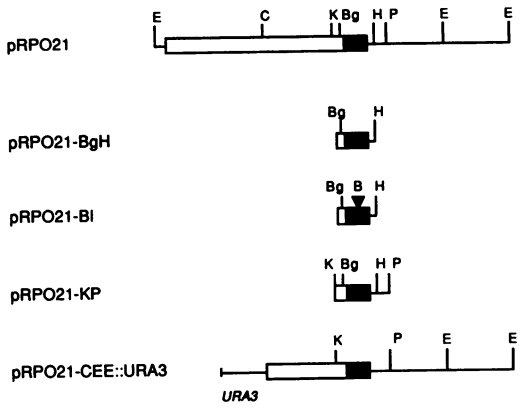

FIG. 1. Plasmid DNAs used in altering the C-terminal domain of the yeast *RPO21* polypeptide. The constructions are described in detail in Materials and Methods. The rectangle in pRPO21 indicates the protein-coding region; the solid portion encodes the C-terminal heptapeptide repeats. Cleavage sites for the restriction endonucleases *Eco*RI (E), *Cla*I (C), *Kpn*I (K), *Bgl*II (Bg), *Hin*dIII (H), *Pst*I (P), and *Bam*HI (B) are shown.

containing either partial *Mbo*I-digested or *Eco*RI-digested Tr-pGA-Amal-5-6 DNA. Mapping, subcloning, and hybridization studies of the DNA inserts of these phages were performed by standard techniques essentially as previously described (17).

DNA sequencing. Restriction fragments of a portion of the hamster DNA insert of λAma E10 and of the *D. melanogaster RPO21* clone p4.1 were subcloned into pEMBL vectors (7). Ordered sets of deletions were made by using an exonuclease III-S1 method (12) or appropriate restriction endonuclease cuts, and single-stranded DNA templates were sequenced by the dideoxy method (24) with either the large fragment of DNA polymerase I or T7 DNA polymerase (Sequenase; U.S. Biochemicals). Both strands of DNA were sequenced. For computer analyses of the nucleotide sequence we used the Cornell DNA sequence package (9), BIONET, and the Protein Identification Resource.

Yeast strains and *RPO21* plasmids. The *S. cerevisiae* diploid strain LP112 (**a**/α *can1-100/can1-100 his3-11,15/his3-11,15 leu2-3,112/leu2-3,112 trp1-1/trp1-1 ura3-1/ura3-1 ade2-1/ade2-1*) was constructed from the isogenic haploid strains W303-1A and W303-1B (23) by S. Lindquist and was obtained from J. Segall. Yeast cells were grown in YEPD medium (1% yeast extract, 2% peptone, 2% glucose) or minimal medium (0.67% yeast nitrogen base without amino acids, supplemented with appropriate amino acids, with 2% glucose). The integrating plasmid pRPO21-CEE::URA3, used to introduce mutant *RPO21* alleles into the chromosomal *RPO21* locus, was constructed from pRPO21, a pEMBL8 construct containing the 7.5-kilobase-pair (kb) and adjacent 1.7-kb *Eco*RI fragments of the *S. cerevisiae RPO21* gene (1a, 17) (Fig. 1). pRPO21 was digested with *Cla*I, deleting nucleotides 1 to 2787 (1a) of the *RPO21* gene and 1 kb of the pEMBL vector. The *Cla*I site was then converted to a unique *Bam*HI site by filling in with the Klenow fragment of DNA polymerase I and ligating with an 8-base-pair (bp) *Bam*HI linker. A *Bam*HI-linkered 1.2-kb fragment of yeast DNA containing the *URA3* gene was cloned at this *Bam*HI site.

To facilitate the deletion of yeast *RPO21* heptapeptide repeats, pRPO21-Bl was constructed from the plasmid pRPO21-BgH by linearizing with *Alu*I and ligating in the presence of an excess of a 10-bp *Bam*HI linker. pRPO21-BgH was one of the exonuclease III deletion constructs used in sequencing the yeast *RPO21* gene (1a) and contains DNA from nucleotides 4675 to 5723. Examination of the resulting plasmids by restriction mapping and DNA sequencing identified pRPO21-Bl, which contains a *Bam*HI linker at nucleotide 5156 in heptapeptide repeat 13. The *RPO21* deletion clones were constructed from pRPO21-Bl by *Bam*HI digestion and exonuclease III and S1 nuclease treatment (12). The size of each deletion in the different isolates was determined by DNA sequencing. The amino acid sequences encoded at the deletion junctions of the DNAs used in this study are indicated in the footnote to Table 1. For each of the deletion constructs, the *Bgl*II-*Hin*dIII fragment with a desired deletion was subcloned into the plasmid pRPO21-KP and then into pRPO21-CEE::URA3 (Fig. 1). The final construct was linearized with *Kpn*I and used to transform spheroplasts of LP112 cells. Chimeric *RPO21* plasmids were constructed by replacing the DNA between the *Hpa*II site at nucleotide 4911 (1a) in the yeast *RPO21* gene and the *Bam*HI site in pRPO21-Bl with *Hpa*II-*Bam*HI fragments obtained from subclones of hamster and *D. melanogaster RPO21* DNA. *Hpa*II sites occur at positions 726 and 581 in the hamster and *D. melanogaster* DNAs, respectively (see Fig. 3A and B). There were *Bam*HI sites in each DNA in a polylinker immediately adjacent to the sequenced 3′ untranslated region of each gene. The chimeric constructs therefore contain all the hamster or *D. melanogaster RPO21* coding sequences distal to the *Hpa*II sites.

RESULTS

Isolation of genomic DNA encoding the hamster *RPO21* polypeptide. The *D. melanogaster RPO21* DNA which encodes the largest subunit of RNA polymerase II (10) has proven to be a valuable probe for the isolation of related DNAs from other eucaryotic species (4, 5, 17). By screening genomic libraries of hamster DNA with the *D. melanogaster* probes p4.2 and p4.1 (generous gifts from the laboratory of A. Greenleaf), we have isolated three λ phage clones with overlapping inserts. These inserts together contain the 7.7- and 10.0-kb *Eco*RI fragments that we had shown previously (16; J. K.-C. Wong, Ph.D. thesis, University of Toronto, Toronto, Ontario, Canada, 1985) encode parts of the hamster RNA polymerase II polypeptide. Restriction mapping and hybridization studies on this hamster DNA are summarized in Fig. 2. These data indicate that the hamster *RPO21* gene spans a minimum of 17 kb of genomic DNA. Since only 6 kb of coding sequence is required to encode the M_r 215,000 *RPO21* polypeptide, this gene must contain introns.

Nucleotide sequences of *D. melanogaster* and hamster *RPO21* DNA. The *RPO21* polypeptides of both yeast (1a) and *D. melanogaster* (2) RNA polymerase II have been shown to be homologous to β′, the largest subunit of the *E. coli* enzyme. In addition, both the yeast (1a) and the analogous mouse (5) *RPO21* polypeptides contain a C-terminal, RNA polymerase II-specific domain composed of a tandemly repeated heptapeptide sequence. To investigate in greater detail both the evolutionary conservation and the function of this domain in *RPO21* polypeptides, we first determined the nucleotide sequence of DNA encoding the carboxy termini of both the hamster and *D. melanogaster RPO21* polypeptides.

Figure 3A shows the hamster *RPO21* nucleotide sequence from a *Pvu*II site within the 3′ portion of the coding region

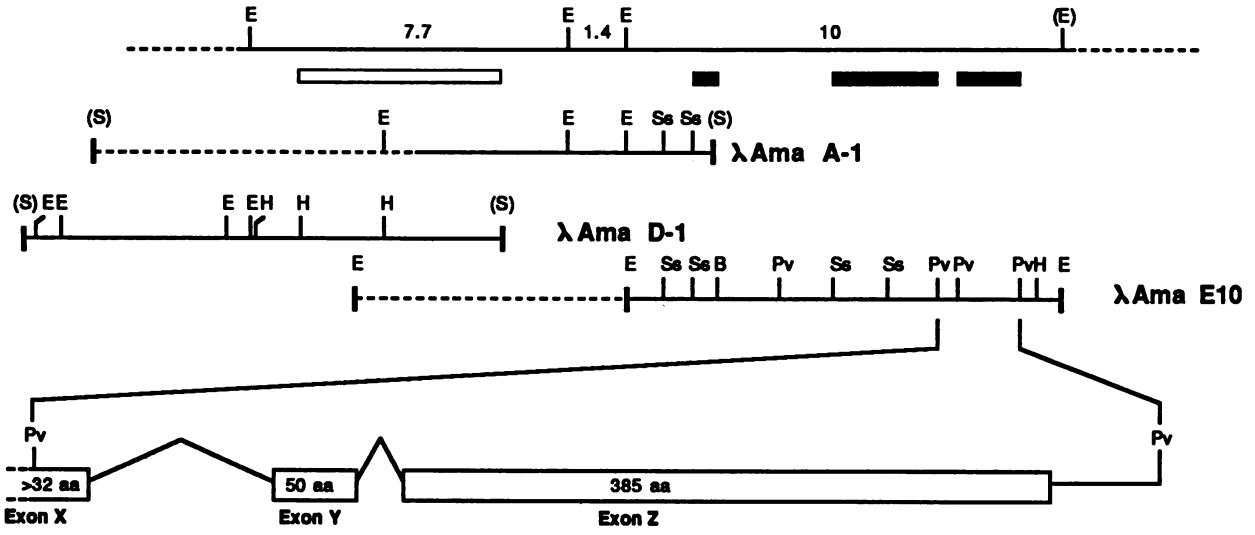

FIG. 2. Restriction maps of the Chinese hamster ovary cell *RPO21* gene and isolated *RPO21* recombinants. Cleavage sites for the restriction endonucleases *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H), *Pvu*II (Pv), *Sst*I (Ss), and *Sal*I (S) on DNA of the inserts of three overlapping *RPO21* recombinant phage clones are shown. Sites in parentheses represent those introduced through cloning steps. Such sites are not in the genomic DNA. Restriction fragments that hybridized to *D. melanogaster RPO21* (*RpII215*) DNA in p4.2 (▭) and p4.1 (▬) are indicated. The lower portion of the figure shows the putative intron-exon organization for the 3′ region of this *RPO21* gene. The inserts of the three clones have not been mapped with the same restriction endonucleases. Both λAma A-1 and λAma E10 are chimeric, each containing chromosomally noncontiguous insert DNAs (-----).

(Fig. 2) to a second *Pvu*II site beyond the stop codon. This sequence encompasses the 3′-most regions of hamster *RPO21* DNA that hybridized to the *D. melanogaster* p4.1 probe (Fig. 2) and encodes the C-terminal 467 amino acids of the hamster *RPO21* polypeptide. Figure 3B shows the nucleotide sequence of the analogous part of the *D. melanogaster RPO21* gene from a *Bam*HI site (coordinate 5.2 in reference 2) to the *Eco*RI site (coordinate 6.8) beyond the stop codon. These 1,746 bp of sequence encode the C-terminal 408 amino acids of the *D. melanogaster RPO21* polypeptide.

The 5′ end of the DNA sequences shown in Fig. 3A and B encodes a stretch of amino acids (underlined in Fig. 3A and B) that is homologous to a region very near the C terminus of the *E. coli* RNA polymerase polypeptide β′. The exon(s) encoding the C-terminal region of these two *RPO21* polypeptides is separated from this procaryoticlike N-terminal region by one or more introns. In the *D. melanogaster* gene the single intron identified in Fig. 3B had been previously mapped to this region by Biggs et al. (2). We have identified two putative C-terminal exons in the hamster gene. The first of these, exon Y in Fig. 2, was identified by the presence of splice donor and acceptor sites and by its amino acid homology with the corresponding part of the *D. melanogaster* polypeptide. We also noted that the nucleotide sequence of the hamster exon Y, as well as its deduced amino acid sequence, was remarkably similar to a portion of sequenced DNA in the mouse *RPO21* gene, which had not been initially identified as an exon by Corden et al. (5, but see also, reference 1). For this putative exon of 150 bp, the mouse and hamster DNAs are 93% homologous and their predicted amino acid sequences are identical. In contrast, the putative intron DNA sequences on either side of exon Y in the hamster and mouse gene show considerably less sequence homology (77 and 67% identity, respectively). Exon Y and its homolog in the *D. melanogaster* polypeptide do not appear to be related to sequences in either the yeast *RPO21* or *RP031* polypeptide, nor is this serine-, glycine-, and proline-rich sequence present in the *E. coli* β′ polypeptide.

The last exon in the hamster gene, exon Z (Fig. 2 to 4), encodes a 385-amino-acid domain. The DNA sequence of this domain, like the 3′ end of the yeast *RPO21* gene (1a) and the analogous mouse gene (5), is composed of a tandemly repeated, somewhat degenerate 21-bp sequence encoding the highly conserved amino acid consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The analogous portion of the *D. melanogaster* gene encodes 316 amino acids. The deduced amino acid sequences of this domain in the yeast, *D. melanogaster*, and hamster polypeptides are shown in Fig. 4. These sequences have been aligned to demonstrate the unusual repetitive nature of this C-terminal domain.

There are 52 repeats of the heptapeptide in the hamster polypeptide. Comparison of the hamster and mouse (5) sequences showed that they are 89% homologous at the nucleic acid level and virtually identical at the amino acid level. There is only a single amino acid difference (a threonine-to-alanine change in repeat 38) between the hamster and mouse sequences. All of the other deviations from the consensus heptapeptide sequence in the hamster domain (shown as boxed residues in Fig. 4), including the single-amino-acid insertion between repeats 2 and 3 and the three-amino-acid insertion between repeats 49 and 50, are perfectly conserved in these two mammalian polypeptides.

The *D. melanogaster* amino acid sequence is considerably more divergent. Although the yeast sequence showed only minimal deviation from the consensus, with 17 of the 26 repeats unchanged, all but two of the somewhat arbitrarily identified 44 *D. melanogaster* repeats contain replacements of one or more of the consensus amino acids. In comparison, the longer mammalian C-terminal domain is more like that of the yeast. The first 26 repeats in the hamster polypeptide are a close match to the consensus, while the distal 26 repeats contain more deviations. These replacements in the hamster polypeptide have a particular pattern. The tyrosine and proline residues are virtually invariant, and a majority of the

A

```
GCT GGC ACT GGC TGT TTT GAC CTC CTG CTC GAC GCT GAA AAG TGC AAA TAT GGC ATG GAG ATC CCC ACC AAT ATC CCT GGC CTG GGG GCT    90
Ala Gly Thr Gly Cys Phe Asp Leu Leu Leu Asp Ala Glu Lys Cys Lys Tyr Gly Met Glu Ile Pro Thr Asn Ile Pro Gly Leu Gly Ala

GCT GGA C  GTGAGTCAGAGGGCTTTGGCTGGGTCTGTGGAGGATGGGTGGGGATAGGCATACGAAGGAGGGTGTCTTTAGTCCCTGAGTGTACTTCCTCAAGAATTACTCTGGGTT   205
Ala Gly A

TCTTTGCAGCTACTGGAATGTTCTTTGGCTCTCACCCAGTCCGATGGGCGGAATATCTCCTGCAATGACACCCTGGAACCAGTCAGACCCAGCCTATGGTGCTTGGTCCCCAAGTGTTG   324

GTGAGTAGCCTAGTCCAGGAAGAGAACTATTGGACTTCAGGGGCTGTGGGGGCTGGGGAAAGCTATGTTGTTGAGCTACAACTGATTCCACCTGTCCTCCTTCTCAG GG AGT GGA   439
                                                                                                              rg Ser Gly

ATG ACC CCA GGA GCA GCT GGC TTC TCG CCC AGT GCT GCA TCT GAT GCC AGT GGC TTT AGC CCA GGT TAC TCT CCT GCA TGG TCT CCG ACA   529
Met Thr Pro Gly Ala Ala Gly Phe Ser Pro Ser Ala Ala Ser Asp Ala Ser Gly Phe Ser Pro Gly Tyr Ser Pro Ala Trp Ser Pro Thr

CCA GGC TCT CCA GGC TCC CCT GGA CCC TCA AGC CCA TAT ATC CCC TCA CCG G   GTGAGTTGCTGTGCTCCTTCCCTCTTCACTCCTTCCTTCTGTGGCAG   628
Pro Gly Ser Pro Gly Ser Pro Gly Pro Ser Ser Pro Tyr Ile Pro Ser Pro G

GCTTCCCATTTCTTATCTCCCTTTTGTCTTTCCCTACAG  GT GGT GCT ATG TCT CCC AGC TAC TCA CCA ACA TCA CCT GCC TAT GAG CCA CGG TCC CCC   726
                                         ly Gly Ala Met Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ala Tyr Glu Pro Arg Ser Pro

GGG GGA TAT ACA CCC CAG AGC CCT TCA TAC TCT CCT ACT TCA CCC TCC TAC TCC CCA ACC TCG CCA TCT TAT TCC CCA ACC AGT CCC AAC   816
Gly Gly Tyr Thr Pro Gln Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Asn

TAT AGC CCT ACG TCA CCT AGC TAC TCG CCG ACC TCT CCA TCC TAC TCG CCA ACC TCT CCA TCC TAC TCC CCG ACC TCT CCA TCC TAC TCC   906
Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser

CCG ACC TCT CCC AGC TAC TCC CCG ACC TCT CCC AGC TAC TCC CCG ACA TCA CCC AGC TAC TCC CCG ACT TCT CCC AGC TAC TCC CCG ACA   996
Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr

TCT CCC AGC TAC TCA CCC ACC TCT CCC AGC TAT TCC CCC ACC TCC CCA AGC TAC TCT CCC ACC TCC CCC AGT TAC TCA CCG ACT TCC CCC  1086
Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro

AGC TAC TCT CCT ACA TCT CCA AGC TAC TCG CCA ACT TCT CCA AGT TAT TCA CCC ACC AGC CCT AAC TAT TCT CCA ACT AGT CCC AAT TAC  1176
Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Asn Tyr Ser Pro Thr Ser Pro Asn Tyr

ACC CCA ACA TCC CCC AGC TAC AGC CCA ACA TCA CCT AGC TAC TCG CCT ACT AGT CCA AAC TAC ACA CCA ACC AGT CCC AAC TAC AGC CCA  1266
Thr Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Asn Tyr Thr Pro Thr Ser Pro Asn Tyr Ser Pro

ACC TCG CCA AGC TAT TCC CCA ACT TCA CCC AGC TAC TCC CCA ACC TCG CCA AGC TAT TCA CCC TCC AGC CCA CGG TAC ACA CCA CAA TCT  1356
Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Ser Ser Pro Arg Tyr Thr Pro Gln Ser

CCC ACC TAC ACA CCA AGC TCT CCC AGC TAC AGC CCC AGC TCA CCA AGC TAC AGC CCC ACC TCA CCC AAG TAT ACC CCA ACT AGT CCT TCT  1446
Pro Thr Tyr Thr Pro Ser Ser Pro Ser Tyr Ser Pro Ser Ser Pro Ser Tyr Ser Pro Thr Ser Pro Lys Tyr Thr Pro Thr Ser Pro Ser

TAC AGT CCC AGC TCT CCA GAG TAT ACC CCA ACC TCT CCC AAA TAC TCA CCT ACA AGC CCT AAA TAT TCA CCT ACT TCT CCC AAG TAT TCT  1536
Tyr Ser Pro Ser Ser Pro Glu Tyr Thr Pro Thr Ser Pro Lys Tyr Ser Pro Thr Ser Pro Lys Tyr Ser Pro Thr Ser Pro Lys Tyr Ser

CCT ACC AGT CCC ACC TAC TCA CCT ACC ACA CCA AAA TAC TCG CCA ACC TCT CCT ACA TAC TCG CCA ACC TCT CCA GTC TAC ACC CCA ACC  1626
Pro Thr Ser Pro Thr Tyr Ser Pro Thr Thr Pro Lys Tyr Ser Pro Thr Ser Pro Thr Tyr Ser Pro Thr Ser Pro Val Tyr Thr Pro Thr

TCC CCC AAG TAC TCA CCC ACC AGC CCC ACC TAC TCC CCC ACC TCT CCC AAG TAC TCA CCC ACC AGC CCC ACC TAC TCC CCC ACC TCC CCC  1716
Ser Pro Lys Tyr Ser Pro Thr Ser Pro Thr Tyr Ser Pro Thr Ser Pro Lys Tyr Ser Pro Thr Ser Pro Thr Tyr Ser Pro Thr Ser Pro

AAG GGC TCC ACT TAC TCC CCC ACC TCC CCT GGT TAC TCA CCT ACT AGC CCC ACC TAC AGC CTC ACC AGC CCA GCC ATC AGC CCT GAT GAC  1806
Lys Gly Ser Thr Tyr Ser Pro Thr Ser Pro Gly Tyr Ser Pro Thr Ser Pro Thr Tyr Ser Leu Thr Ser Pro Ala Ile Ser Pro Asp Asp

AGT GAT GAG GAA AAC TAA GGGTGAGCAGGGCGCAGCAGCAGCAGGTTAGGGTCAGACAGCCTTAGTGGCCTGTGCAGTCACTTCCCTTGAGCTGTGACCCTAGCCTGGGCCCC  1919
Ser Asp Glu Glu Asn  .

TTGTACATAACTCCTTGTGACAGAACCCTCTGGAGGTTCTAGATCCCATTTTTGATGGGCTTTTCGTCTTGTCCTCACTCGTGCTGTCTTGGGACTCACTGACAGCTG          2027
```

changes in the distal hamster repeats occur at position 7 (as arbitrarily aligned in Fig. 4). In many repeats the consensus serine residue at this position is replaced, most frequently by a charged amino acid residue such as lysine.

Deletions of heptapeptide repeats in the yeast *RPO21* polypeptide. The remarkable evolutionary conservation of the *RPO21* C-terminal heptapeptide repeat domain shown in this study and its presence only in an RNA polymerase II polypeptide suggest that it must play an important and polymerase II-specific role in transcription. To initiate a genetic assessment of structure and function relationships of this novel repeating heptapeptide domain, we constructed a series of plasmid vectors that permit replacement of the wild-type yeast chromosomal *RPO21* DNA sequences, encoding the C terminus of this polypeptide, with altered sequences. For these studies we first altered the yeast *RPO21* DNA within the plasmid pRPO21-Bl (Fig. 1), verified these alterations by DNA sequencing, and then sequentially transferred these altered DNAs into the plasmid pRPO21-KP and then into pRPO21-CEE::URA3 (see Materials and Methods). Transformation of *ura3*$^-$/*ura3*$^-$ yeast strains with pRPO21-CEE::URA3, linearized at its unique *Kpn*I site, targeted integration (22) of this DNA into the *RPO21* locus on chromosome IV (13). The integration generated a partial duplication of *RPO21*; the intact copy of the gene contains the 3′ deletion; the other copy has a 5′ deletion of promoter and coding sequences and is therefore nonfunctional.

To first demonstrate that the heptapeptide repeats in the *S. cerevisiae RPO21* polypeptide do indeed serve an essential function in the yeast RNA polymerase II enzyme, we made a series of constructs which contained progressively larger deletions within the DNA encoding the yeast repeats. All of these deletion constructs left intact the short stretch of nonconsensus amino acids at the extreme C terminus of this *RPO21* polypeptide (see Materials and Methods). Using the *URA3* integrating plasmid vector pRPO21-CEE::URA3, we transformed diploid yeast cells with these deletion constructs. Restriction enzyme digestions and Southern blot

B

```
GAT CCC ATG AGG GGC GTC TCT GAG AAC ATT ATC ATG GGA CAG CTG CCC AAA ATG GGT ACC GGC TGC TTT GAC CTT CTG CTC GAT GCA GAG   90
Asp Pro Met Arg Gly Val Ser Glu Asn Ile Ile Met Gly Gln Leu Pro Lys Met Gly Thr Gly Cys Phe Asp Leu Leu Leu Asp Ala Glu

AAG TGT CGT TTC GGC ATC GAG ATT CCC AAT ACG TTG GGC AAC A   GTATGCTAGGTGGCGCCGCTATGTTCATTGGCGGTGGATCGACACCGAGCATGACGCCA  192
Lys Cys Arg Phe Gly Ile Glu Ile Pro Asn Thr Leu Gly Asn I

CCGATGACGCCGTGGGCTAACTGCAACACGCCGCGATACTTCTCTCCACCCGGCCACGGTAAGTAAATCTATCCATTATGGGTTTTTGATTTTACTCTCGACTCAGCAACGCTATATCT  311

TTTGTTAATTAACTTTTTTGATGCGATTGCCTAATATATTCAATCTCTTTCTTTCTTGTGCTTGCAG  TA AGT GCC ATG ACT CCT GGC GGT CCC AGT TTC TCG CCT  416
                                                                      le Ser Ala Met Thr Pro Gly Gly Pro Ser Phe Ser Pro

TCG GCT GCA TCG GAT GCG TCC GGA ATG TCG CCT AGC TGG TCG CCG GCT CAT CCG GGC TCA TCG CCC AGT TCA CCA GGA CCT TCG ATG TCG  506
Ser Ala Ala Ser Asp Ala Ser Gly Met Ser Pro Ser Trp Ser Pro Ala His Pro Gly Ser Ser Pro Ser Ser Pro Gly Pro Ser Met Ser

CCG TAT TTC CCA GCC TCG CCG AGT GTT TCT CCC TCT TAT TCG CCA ACG AGT CCG AAC TAC ACG GCA TCT TCT CCC GGT GGA GCC TCG CCG  596
Pro Tyr Phe Pro Ala Ser Pro Ser Val Ser Pro Ser Tyr Ser Pro Thr Ser Pro Asn Tyr Thr Ala Ser Ser Pro Gly Gly Ala Ser Pro

AAT TAC TCG CCC TCG AGT CCG AAC TAT TCG CCG ACG TCG CCG CTT TAT GCA AGT CCA CGT TAC GCA TCG ACA ACG CCA AAT TTC AAT CCA  686
Asn Tyr Ser Pro Ser Ser Pro Asn Tyr Ser Pro Thr Ser Pro Leu Tyr Ala Ser Pro Arg Tyr Ala Ser Thr Thr Pro Asn Phe Asn Pro

CAG TCG ACG GGT TAC TCG CCA TCT TCA TCG GGA TAC TCG CCA ACA TCC CCG GTC TAC TCG CCC ACG GTG CAA TTC CAG TCG AGT CCG TCG  776
Gln Ser Thr Gly Tyr Ser Pro Ser Ser Ser Gly Tyr Ser Pro Thr Ser Pro Val Tyr Ser Pro Thr Val Gln Phe Gln Ser Ser Pro Ser

TTT GCG GGC AGC GGT AGC AAC ATT TAC TCG CCG GGC AAT GCG TAC TCG CCC AGC TCG TCC AAC TAC TCC CCC AAT TCA CCA TCC TAC TCG  866
Phe Ala Gly Ser Gly Ser Asn Ile Tyr Ser Pro Gly Asn Ala Tyr Ser Pro Ser Ser Ser Asn Tyr Ser Pro Asn Ser Pro Ser Tyr Ser

CCG ACA TCA CCA TCG TAC TCG CCG TCA AGT CCT TCG TAC TCG CCA ACG TCG CCT TGC TAT TCG CCC ACA TCG CCT TCG TAC TCG CCA ACG  956
Pro Thr Ser Pro Ser Tyr Ser Pro Ser Ser Pro Ser Tyr Ser Pro Thr Ser Pro Cys Tyr Ser Pro Thr Ser Pro Ser Tyr Ser Pro Thr

AGT CCG AAC TAC ACA CCC GTA ACA CCC TCA TAC TCG CCG ACA AGT CCG AAC TAT TCA GCG TCG CCG CAA TAT TCT CCA GCC TCG CCA GCT 1046
Ser Pro Asn Tyr Thr Pro Val Thr Pro Ser Tyr Ser Pro Thr Ser Pro Asn Tyr Ser Ala Ser Pro Gln Tyr Ser Pro Ala Ser Pro Ala

TAC TCG CAA ACG GGG GTG AAG TAC TCA CCG ACA TCG CCG ACG TAC TCG CCG CCG TCA CCA TCG TAC GAT GGG TCT CCC GGA TCA CCA CAA 1136
Tyr Ser Gln Thr Gly Val Lys Tyr Ser Pro Thr Ser Pro Thr Tyr Ser Pro Pro Ser Pro Ser Tyr Asp Gly Ser Pro Gly Ser Pro Gln

TAT ACG CCA GGA TCT CCG CAG TAC TCG CCG GCC TCG CCT AAG TAC TCG CCG ACC TCA CCG CTG TAC TCG CCC AGC TCG CCG CAG CAC TCG 1226
Tyr Thr Pro Gly Ser Pro Gln Tyr Ser Pro Ala Ser Pro Lys Tyr Ser Pro Thr Ser Pro Leu Tyr Ser Pro Ser Ser Pro Gln His Ser

CCC TCA AAC CAG TAC AGC CCA ACA GGA TCG ACC TAT TCG GCG ACG AGT CCG CGG TAC TCG CCG AAC ATG TCC ATC TAC TCG CCG AGC AGC 1316
Pro Ser Asn Gln Tyr Ser Pro Thr Gly Ser Thr Tyr Ser Ala Thr Ser Pro Arg Tyr Ser Pro Asn Met Ser Ile Tyr Ser Pro Ser Ser

ACC AAG TAC TCG CCC ACC TCG CCA ACG TAC ACA CCG ACG GCC CGC AAC TAC TCG CCC ACG TCA CCG ATG TAC TCG CCA ACG GCT CCA TCG 1406
Thr Lys Tyr Ser Pro Thr Ser Pro Thr Tyr Thr Pro Thr Ala Arg Asn Tyr Ser Pro Thr Ser Pro Met Tyr Ser Pro Thr Ala Pro Ser

CAC TAC AGT CCC ACG AGT CCG GCC TAC TCG CCC AGC AGT CCC ACG TTC GAG GAG AGC GAA GAC TGA GGAAGGGAGGACGGGGGTAGCTCCCCCAGCA 1503
His Tyr Ser Pro Thr Ser Pro Ala Tyr Ser Pro Ser Ser Pro Thr Phe Glu Glu Ser Glu Asp  .

CGTCCCGATCTCCGGGGCGGTCAAGCCGTAGTTAAGTACCGACTACGCTCGATCGCCCGAACGGGAAAGGTGAATTTAATAATATTATGTTTGTATACAGCAAATAGGTCTCTCATACA 1622

AAATTCAATACTGTTTGCAAAAGTAGCTTTAAACAAGTTGTATAAATTAGCCTACAGTTTCGTATTTGGTTTTGTGGTAAAGTAAACATTTTGAATTGTTTGCGTTAACCAGAAAAGCG 1741

AATTC                                                                                                                  1746
```

FIG. 3. Nucleotide and predicted amino acid sequences at the 3′ end of the Chinese hamster ovary cell (A) and the *D. melanogaster* (B) *RPO21* gene. The DNA sequence between two *Pvu*II sites (Fig. 1) in the hamster *RPO21* DNA of λAma E10 and between the *Bam*HI and *Eco*RI sites (coordinates 5.2 to 6.8 in reference 2) in the *D. melanogaster* DNA was determined. The nucleotides are numbered from the *Pvu*II or the *Bam*HI site in the 5′-to-3′ direction. The location of putative exons was deduced not only from the presence of consensus splice donor and acceptor sites but also from the homology of these amino acid sequences to corresponding regions of yeast *RPO21* DNA (1a) and the analogous mouse gene (1, 8). Amino acids identical to those present near the C terminus of the *E. coli* β′ polypeptide are underlined. The *Hpa*II sites, used in constructing the chimeric yeast-hamster and yeast-*D. melanogaster* genes, are boxed.

experiments (data not shown) with the genomic DNA of the Ura$^+$ transformants were used to verify that one of the two chromosomal *RPO21* genes had in each case been appropriately altered.

The analysis of the resulting *RPO21*$^+$/*rpo21*-Δ::*URA3* diploid strains is summarized in Table 1. Viable heterozygous diploid (wild-type/deletion) strains were obtained with each deletion. All of the deletions had either a recessive phenotype or no detectable phenotype at all. The two largest deletions which removed either 123 or 119 amino acids within the heptapeptide repeats were lethal in haploid strains. After sporulation of these two *RPO21*$^+$/*rpo21*-Δ::*URA3* strains, the resulting tetrads contained only two viable spores. Deletions which removed 104, 88, 61, or 52 amino acids from within the heptapeptide repeats, however, gave rise to viable haploid strains. The tetrads in each case contained four viable spores in which the Ura$^+$ phenotype segregated 2:2. These data indicate that a minimum number of the heptapeptide repeats, between 9 and 11, is essential for *RPO21* function and yeast viability. The haploid strains containing an *RPO21* polypeptide with either 104 or 88 amino acids deleted from the C-terminal domain grew more slowly than did the wild-type strains at 30°C. Our preliminary characterization of these two deletion mutants has indicated that they are cold sensitive at 21°C (data not shown).

Species specificity of the *RPO21* heptapeptide repeat domain. Although the heptapeptide repeats are evolutionarily conserved in the *RPO21* polypeptides of apparently all eucaryotes, our DNA sequencing studies have revealed that significant differences from the consensus sequence occur in the yeast, *D. melanogaster*, and hamster *RPO21* polypeptides. In addition, there is a striking difference in the length of this domain in different species. To assess the significance of these species-specific differences, we have replaced the yeast *RPO21* C-terminal heptapeptide repeats with the more divergent *D. melanogaster* or the much longer mammalian domain.

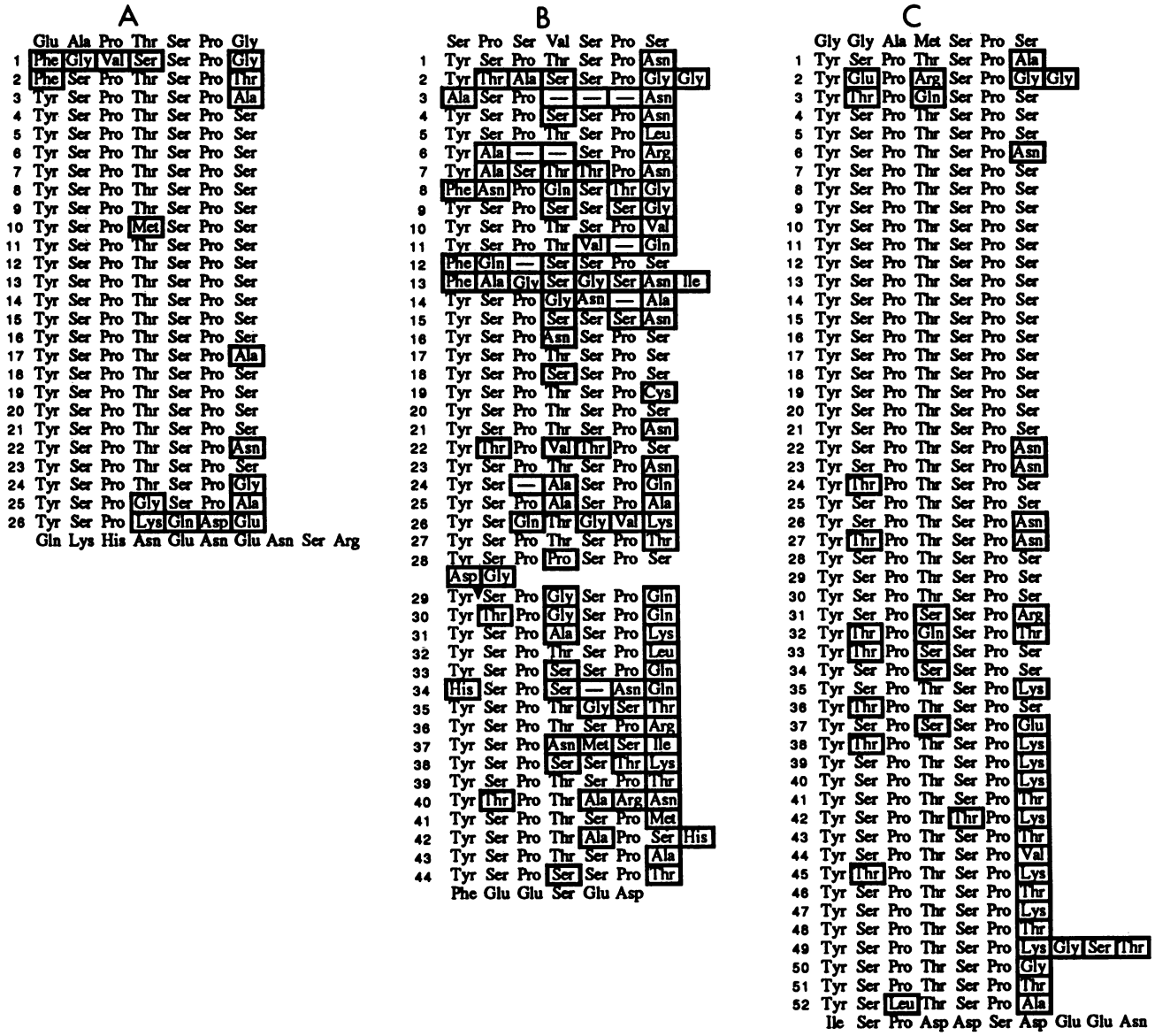

FIG. 4. Amino acid sequences of the C-terminal domains of the *S. cerevisiae* (1a) (A), *D. melanogaster* (B), and hamster (C) *RPO21* polypeptides. The boxed residues indicate amino acids within the heptapeptide repeats that differ from the consensus heptapeptide sequence.

Construction of such chimeric *RPO21* strains of *S. cerevisiae* was facilitated by the presence of a conserved *Hpa*II site in the DNAs encoding the C-terminal domains of all three polypeptides. This *Hpa*II site occurs at a Pro-Gly sequence found at a similar place near the beginning of the heptapeptide repeats of all three species. Constructions which precisely exchange DNA encoding the distal 25 repeats of the yeast *RPO21* polypeptide with DNA encoding either heptapeptide repeats 3 to 44 of the *D. melanogaster* or repeats 3 to 52 of the hamster *RPO21* polypeptide were first made in the plasmid pRPO21-Bl (see Materials and Methods). Subsequent transfer of these altered sequences into the integrating plasmid vector pRPO21-CEE::URA3 and integration into the yeast chromosomal *RPO21* gene permitted an in vivo assay of chimeric *RPO21* function.

As with the deletion of the yeast *RPO21* sequences described above, we again used the transformation of diploid yeast cells, sporulation, and tetrad dissection to determine the phenotype conferred by these chimeric *RPO21* polypeptides. The diploid yeast strain *RPO21*$^+$/*rpo21-Dm*::*URA3*, with chimeric yeast-*D. melanogaster* *RPO21* DNA integrated at one of its two *RPO21* loci, produced two viable spores and two dead spores. All the living spores had a Ura$^-$ phenotype and had an unaltered wild-type *RPO21* gene. Thus, not only do the yeast *RPO21* heptapeptide repeats seem essential, but also they cannot be replaced by DNA encoding the more divergent *D. melanogaster* C-terminal domain.

In distinct contrast to this result, however, was our finding that a similar replacement of the yeast *RPO21* domain by the analogous hamster C-terminal domain resulted in no detectable mutant phenotype. Sporulation of the diploid transformant *RPO21*$^+$/*RPO21-ha*::*URA3* produced four viable spores in which the Ura$^+$ phenotype segregated 2:2 (Table 1). In further tests of a resulting haploid *RPO21-ha*::*URA3* strain containing only the chimeric yeast-hamster *RPO21*

TABLE 1. Tetrad analysis of mutant *RPO21/rpo21::URA3* diploid yeast strains

Genotype	Repeats remaining	No. of viable spores per tetrad				Ratio of Ura$^+$ to Ura$^-$ spores
		4	3	2	1	
RPO21 deletions[a]						
RPO21/rpo21-Δ123::URA3	8 3/7	0	0	11	2	0:24
RPO21/rpo21-Δ119::URA3	9	0	0	13	1	1:26
RPO21/rpo21-Δ104::URA3	11 1/7	11	1	2	0	26:25
RPO21/rpo21-Δ88::URA3	13 3/7	11	3	0	0	26:27
RPO21/rpo21-Δ61::URA3	17 2/7	13	1	0	0	27:29
RPO21/rpo21-Δ52::URA3	18 4/7	10	4	0	1	28:25
RPO21 chimeras						
RPO21/rpo21-ha::URA3		15	1	0	0	31:32
RPO21/rpo21-Dm::URA3		0	0	13	6	0:32

[a] The amino acids at each of the deletion junctions are as follows: Δ123, repeat 1-FGVSSP/TSPS-repeat 19; Δ119, repeat 2-FSPTSPT/YSPTSPS-repeat 20; Δ104, repeat 4-TSR/PTSPS-repeat 19; Δ88, repeat 4-YSPTSPS/SPA-repeat 17; Δ61, repeat 6-YSPT/PTSPS-repeat 15; Δ52, repeat 7-YSPTS/SPTSPS-repeat 15.

gene, we found no significant difference in the growth rates at 30°C between an *RPO21*$^+$ strain and the *RPO21-ha::URA3* strain. Furthermore, homozygous diploid strains of this yeast-hamster *RPO21* chimera were able to sporulate efficiently (data not shown). Thus, it appears that the much longer mammalian C-terminal domain can substitute for the functions of the shorter *S. cerevisiae* domain in a yeast RNA polymerase II enzyme.

DISCUSSION

Conserved structures in procaryotic and eucaryotic RNA polymerases. RNA polymerase structure has been exceptionally well conserved in a wide variety of organisms. The homology between procaryotic and eucaryotic RNA polymerases was first noted in our comparison of the deduced amino acid sequences of the yeast *RPO21* and *RPO31* polypeptides and the sequence of the *E. coli* RNA polymerase polypeptide β′ (1a). It has since been extended to the *RPO21* polypeptides of both *D. melanogaster* (2) and mice (1) and to a vaccinia virus-encoded RNA polymerase (3). Furthermore, the conservation of RNA polymerase polypeptide structure between procaryotes and eucaryotes is not restricted to the largest polypeptide. Sweetser et al. (26) have recently shown that the second-largest (M_r 138,750) subunit of yeast RNA polymerase II is homologous to the β subunit of *E. coli* RNA polymerase. Although the nucleotide sequencing studies reported here have not focused upon the β′-like regions of the hamster and *D. melanogaster RPO21* polypeptides, the limited sequence homologies shown in Fig. 3 of this study, when taken together with a partial nucleotide sequence already reported for the N terminus of the *D. melanogaster RPO21* gene (2) and the other reported homologies, do further emphasize this exceptional conservation of structure. Common structural features in these eucaryotic and procaryotic polypeptides must be necessary for the polymerization activity of these enzymes. In addition to the conserved structures in the core subunits of these enzymes, accessory factors involved in the initiation or termination of transcription in procaryotic systems may have their analogs in eucaryotic transcription complexes.

The C-terminal domain of the *RPO21* polypeptides. The carboxyl terminus of the largest subunit of eucaryotic RNA polymerase II has very unusual structural features. In both the hamster and *D. melanogaster* DNA, this C terminus is encoded by separate exons. The domain encoded by the penultimate exon (exon Y in Fig. 2) of the hamster gene appears to be found only in the *RPO21* polypeptides of higher eucaryotes. A very similar serine-, proline-, and glycine-rich sequence is also present at a corresponding position in the *D. melanogaster RPO21* polypeptide. Although the yeast *RPO21* polypeptide has a stretch of about 60 amino acids between the C terminus of the procaryotic-like domain and the beginning of the tandemly repeated heptapeptide sequence, the sequence of this stretch is not homologous to that present in the hamster or *D. melanogaster* polypeptides.

The domain encoded by exon Z in the hamster gene and by most of the last exon of the *D. melanogaster* gene consists of the heptapeptide repeats. This sequence appears to be present in the *RPO21* polypeptides of all eucaryotes but is not present in the analogous RNA polymerase III polypeptide (1a) or in the β′ subunit of *E. coli* RNA polymerase. The remarkable evolutionary conservation of these heptapeptide repeats (Fig. 4) suggests that they provide an important domain that serves a special function associated with the synthesis of mRNA by RNA polymerase II. In this study we have used the techniques of gene replacement in *S. cerevisiae* to directly demonstrate the importance of this C-terminal domain for RNA polymerase II function in vivo.

The deletion mutants that we have constructed to assess the requirement for heptapeptide repeats in the yeast *RPO21* polypeptide were made by removing DNA encoding internal heptapeptide repeats rather than by making a simple truncation of the *RPO21* polypeptide. The yeast, *D. melanogaster*, and hamster *RPO21* polypeptides all have an additional stretch of 6 to 14 amino acids at their extreme C termini distal to the heptapeptide repeats. The yeast *RPO21* deletion mutations we describe in this study therefore addressed the requirement for heptapeptide repeats independent of any additional requirement for the extreme C-terminal amino acids. The deletions of *RPO21* DNA sequences, which result in mutant *RPO21* alleles lacking 123 or 119 amino acids within the heptapeptide repeats, encode *RPO21* polypeptides with only 8 3/7 or 9 repeats, respectively. These two *RPO21* deletions are recessive lethal mutations. In contrast, deletions of 104 or fewer amino acids within this part of the *RPO21* polypeptide are compatible with yeast viability. These deletions thus define a minimum requirement for between 9 and 11 repeats in the yeast *RPO21* polypeptide. A similar requirement for about one-half of the 52 repeats in the mouse polypeptide for *RPO21* function in mammalian cells has also been demonstrated recently (1b).

The *D. melanogaster* and hamster *RPO21* polypeptides have an analogous C-terminal domain. By constructing chimeric yeast *RPO21* strains with *D. melanogaster* or hamster DNA replacing the yeast C-terminal coding region,

we have demonstrated a species specificity to the functioning of this part of the RNA polymerase II polypeptide. The 44 *D. melanogaster* heptapeptide repeats are considerably more divergent than those in either the yeast polypeptide or the 26 proximal mammalian repeats. For this reason alone, it is not surprising that the yeast-*D. melanogaster* chimeric construct was a recessive lethal mutation. Lack of function of this particular chimeric gene could result from loss of enzyme activity, instability of the chimeric polypeptide, or inefficient assembly of this polypeptide into the multisubunit RNA polymerase II enzyme. Nevertheless, our observation that the analogous replacement of yeast DNA by DNA encoding the C terminus of the hamster *RPO21* polypeptide resulted in viable haploid yeast strains permits certain conclusions. First, the nonconsensus amino acids at the very C terminus of the yeast polypeptide are not required for any special function in yeast cells that cannot be provided by the different amino acids present in the hamster polypeptide. Second, the replacements of consensus amino acids in the yeast polypeptide that occur in repeats 2, 3, 10, and 17 appear to be dispensable deviations from the conserved consensus heptapeptide sequence. None of these replacements of consensus amino acids, except for a serine-to-asparagine change in repeat 22, are present in the hamster domain. The most striking common feature of the yeast and mammalian amino acid sequences that is absent from the *D. melanogaster* domain is a large region of nearly perfect adherence to the consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser sequence. It may be that this feature alone is essential for *RPO21* function in yeast cells.

A proposed function for the C-terminal *RPO21* heptapeptides. A role for the RNA polymerase II heptapeptide repeat domain in the initiation of transcription is suggested by some indirect experiments. The proteolysis of the largest subunit of RNA polymerase II, which occurs on purification of the enzyme, removes this C-terminal domain (1a, 5). A monoclonal antibody, now believed to recognize this domain, has been shown to block accurately initiated in vitro transcription (6). Furthermore, a synthetic peptide analog containing five of the heptapeptide repeats inhibits the formation of a stable initiation-competent complex at the adenovirus major late promoter in vitro (M. Moyle and C. J. Ingles, unpublished observations). This C-terminal domain may therefore play a special role in the process, which leads to correct initiation of transcription by polymerase II at bona fide promoters. Our observations, reported here, that the domain is essential for *RPO21* function in yeast cells are consistent with a role for the heptapeptide repeats in the processes of transcription initiation.

Several possibilities for a function of this unusual *RPO21* domain have been suggested (1a, 5, 18). The heptapeptide sequence might bind DNA; it might interact with a structural element in the nucleus; it might, if phosphorylated, bind histones and so destabilize nucleosomes (5). However, we find the suggestion (18) that it is involved in binding a special class of RNA polymerase II transcription factors especially attractive. A tandemly repeated structure at the C terminus of the *RPO21* polypeptide could provide a set of sites for interaction with DNA-binding proteins that are present in RNA polymerase II initiation complexes at variable (Fig. 5A) or multiple (Fig. 5B) positions relative to polymerase II located at the start site of transcription. For most eucaryotic genes, one such DNA-binding transcription factor is the protein that recognizes the TATA boxes. Other transcription factors that bind upstream activating sequences or enhancers may also directly contact this domain of RNA polymerase II. If so, we might expect these transcription factors to have a common structural feature for interaction with the heptapeptide repeats. Indeed, short segments rich in acidic amino acid residues are present in several transcriptional activators and are required for positive activation (14, 19). These acidic amino acids could be involved in forming a network of hydrogen bonds between the transcription factor and hydroxyl groups on the side chains of amino acids on the *RPO21* C-terminal repeats. Since the promoters of eucaryotic genes appear to be a combinatorial array of different protein-binding DNA sequences, it is possible that the relative strengths of promoters in different tissues or in cells in different physiological states are influenced by the degree to which different *trans*-acting factors have competing, overlapping, or separate recognition sites on the *RPO21* C-terminal domain.

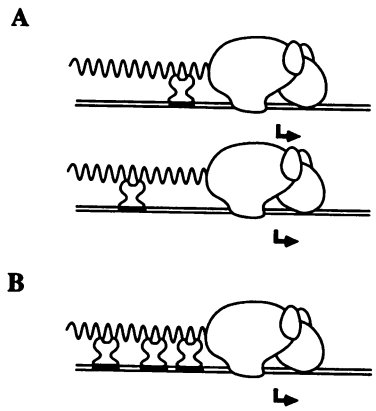

FIG. 5. Model for the function of the tandemly repeated *RPO21* C-terminal domain. RNA polymerase II, shown with two large and several smaller subunits, is positioned at the start site of transcription of a gene. A DNA-binding protein is shown located at variable (A) or multiple (B) positions relative to polymerase II. This DNA-binding protein is bifunctional, contacting its cognate DNA sequence with one domain and interacting with the *RPO21* tandemly repeated heptapeptide sequence with another. The structure of the C-terminal domain of *RPO21* is shown here as a hypothetical extended structure. Its actual structure remains to be determined.

An element of flexibility for these C-terminal heptapeptide repeats may be contributed by the amino acid sequence encoded by exon Y of the hamster *RPO21* gene. This unusual proline-, glycine-, and serine-rich sequence appears to be present only in the *RPO21* polypeptide of higher eucaryotes such as *D. melanogaster* and mammals. It may provide a flexible hinge region between the part of the *RPO21* polypeptide that is homologous to the procaryotic β′ subunit and the tandem heptapeptide repeats. Perhaps its presence only in higher eucaryotes accounts in part for an additional degree of flexibility in positioning upstream activating sequence and enhancer elements.

Our current studies are aimed at demonstrating, by both biochemical and genetic means, direct contacts between this C-terminal domain of the *RPO21* polypeptide and polymerase II transcription factors. It will prove interesting indeed if this unusual domain in RNA polymerase II provides the link between proximal and distal *cis*-acting promoter sequences, the sets of transcription factors that recognize these DNA sequence elements, and eucaryotic RNA polymerase II.

ACKNOWLEDGMENTS

We thank Lina Demirjian and Michael Shales for technical assistance, Jack Greenblatt, Jacqueline Segall, and Valerie Watt for

helpful discussion and suggestions, and A. Greenleaf for his gift of *D. melanogaster RPO21* DNA.

This work was supported by grants MT 4649 and MT 7912 from the Medical Research Council (Canada), a grant from the National Cancer Institute of Canada, and grant 1U41 RR01685-01 from the National Institute of Health to the Bionet Resource. J.K.-C.W., L.A.A., and V.D.F. held Medical Research Council (Canada) Studentships.

ADDENDUM IN PROOF

A recent study of *RP021* deletion mutations (M. Nonet, D. Sweetser, and R. A. Young, Cell **50**:909–915, 1987) also showed that a minimum of 10 heptapeptide repeats is required for *RP021* function in *S. cerevisiae*.

LITERATURE CITED

1. **Ahearn, J. M., Jr., M. S. Bartolomei, M. L. West, L. J. Cisek, and J. L. Corden.** 1987. Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. J. Biol. Chem. **262**:10695–10705.

1a. **Allison, L. A., M. Moyle, M. Shales, and C. J. Ingles.** 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell **42**:599–610.

1b. **Bartolomei, M. S., N. F. Halden, C. R. Cullen, and J. L. Corden.** 1988. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. Mol. Cell. Biol. **8**:330–339.

2. **Biggs, J., L. L. Searles, and A. L. Greenleaf.** 1985. Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II. Cell **42**:611–621.

3. **Broyles, S. S., and B. Moss.** 1986. Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits. Proc. Natl. Acad. Sci. USA **83**:3141–3145.

4. **Cho, K. W. Y., K. Khalili, R. Zandomeni, and R. Weinmann.** 1985. The gene encoding the largest subunit of human RNA polymerase II. J. Biol. Chem. **260**:15204–15210.

5. **Corden, J. L., D. L. Cadena, J. M. Ahearn, Jr., and M. E. Dahmus.** 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc. Natl. Acad. Sci. USA **82**:7934–7938.

6. **Dahmus, M. E., and C. Kedinger.** 1983. Transcription of adenovirus-2 major late promoter inhibited by monoclonal antibody directed against RNA polymerase IIo and IIA. J. Biol. Chem. **258**:2303–2307.

7. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**:1645–1655.

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

9. **Fritensky, B., J. Lis, and R. Wu.** 1981. Portable microcomputer software for nucleotide sequence analysis. Nucleic Acids Res. **10**:6451–6463.

10. **Greenleaf, A. L.** 1983. Amanitin-resistant RNA polymerase II mutations are in the enzyme's largest subunit. J. Biol. Chem. **258**:13403–13406.

11. **Greenleaf, A. L., J. R. Weeks, R. A. Voelker, S. Ohnishi, and B. Dickson.** 1980. Genetic and biochemical characterization of mutants at an RNA polymerase II locus in D. melanogaster. Cell **21**:785–792.

12. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.

13. **Himmelfarb, H. J., E. M. Simpson, and J. D. Friesen.** 1987. Isolation and characterization of temperature-sensitive RNA polymerase II mutants of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **7**:2155–2164.

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

15. **Ingles, C. J.** 1978. Temperature-sensitive RNA polymerase II mutations in Chinese hamster ovary cells. Proc. Natl. Acad. Sci. USA **75**:405–409.

16. **Ingles, C. J., J. Biggs, J. K.-C. Wong, J. R. Weeks, and A. L. Greenleaf.** 1983. Identification of a structural gene for an RNA polymerase II polypeptide in *Drosophila melanogaster* and mammalian species. Proc. Natl. Acad. Sci. USA **80**:3396–3400.

17. **Ingles, C. J., H. J. Himmelfarb, M. Shales, A. L. Greenleaf, and J. D. Friesen.** 1984. Identification, molecular cloning, and mutagenesis of *Saccharomyces cerevisiae* RNA polymerase genes. Proc. Natl. Acad. Sci. USA **81**:2157–2161.

18. **Ingles, C. J., M. Moyle, L. A. Allison, J. K.-C. Wong, J. Archambault, and J. D. Friesen.** 1987. Molecular genetics of eukaryotic RNA polymerases. UCLA Symp. Mol. Cell. Biol. **52**:383–393.

19. **Ma, J., and M. Ptashne.** 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell **48**:847–853.

20. **Mortin, M. A., and G. Lefevre.** 1981. An RNA polymerase II mutation in *Drosophila melanogaster* that mimics ultrabithorax. Chromosoma **82**:237–247.

21. **Moyle, M., T. Hofmann, and C. J. Ingles.** 1986. The *RPO31* gene of *Saccharomyces cerevisiae* encodes the largest subunit of RNA polymerase III. Biochem. Cell Biol. **64**:717–721.

22. **Orr-Weaver, T. L., J. W. Szostak, and R. J. Rothstein.** 1981. Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. USA **78**:6354–6358.

23. **Petko, L., and S. Lindquist.** 1986. Hsp26 is not required for growth at high temperatures, nor for thermotolerance, spore development, or germination. Cell **45**:885–894.

24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

25. **Searles, L. L., R. S. Jokerst, P. M. Bingham, R. A. Voelker, and A. L. Greenleaf.** 1982. Molecular cloning of sequences from a Drosophila RNA polymerase II locus by P element transposon tagging. Cell **31**:585–592.

26. **Sweetser, D., M. Nonet, and R. A. Young.** 1987. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc. Natl. Acad. Sci. USA **84**:1192–1196.